31761 06705603 6

UNIV. OF
TORONTO
LIBRARY

Digitized by the Internet Archive
in 2007 with funding from
Microsoft Corporation

http://www.archive.org/details/mathematicsofrel00rainuoft

Phy
R.

MATHEMATICS OF RELATIVITY

LECTURE NOTES

BY

G. Y. RAINICH

ASSOCIATE PROFESSOR OF MATHEMATICS
UNIVERSITY OF MICHIGAN

(All Rights Reserved)
(Printed in U. S. A.)

28699
8:5:33

EDWARDS BROTHERS, INC.
Lithoprinters and Publishers
ANN ARBOR, MICHIGAN

C O N T E N T S

Introduction

Page

Chapter I. OLD PHYSICS 1
1. Motion of a Particle. The Inverse Square Law..... 1
2. Two Pictures of Matter 3
3. Vectors, Tensors, Operations 5
4. Maxwell's Equations 7
5. The Stress-Energy Tensor 9
6. General Equations of Motion. The Complete Tensor 10

Chapter II. NEW GEOMETRY 12
7. Analytic Geometry of Four Dimensions 12
8. Axioms of Four-Dimensional Geometry 14
9. Tensor Analysis 16
10. Complications Resulting From Imaginary Coordinate..... 20
11. Are the Equations of Physics Invariant 22
12. Curves in the New Geometry 24

Chapter III. SPECIAL RELATIVITY 26
13. Equations of Motion 26
14. Lorentz Transformations 28
15. Addition of Velocities 29
16. Light Corpuscles, or Photons 31
17. Electricity and Magnetism in Special Relativity 33

Chapter IV. CURVED SPACE 35
18. Curvature of Curves and Surfaces 35
19. Generalizations 37
20. The Riemann Tensor 39
21. Vectors in General Coordinates 41
22. Tensors in General Coordinates 44
23. Covariant and Contravariant Components 46
24. Physical Coordinates as General Coordinates 49
25. Curvilinear Coordinates in Curved Space 50
26. New Derivation of the Riemann Tensor 53
27. Differential Relations for the Riemann Tensor 54
28. Geodesics..... 55

Chapter V. GENERAL RELATIVITY 57
29. The Law of Geodesics 57
30. Solar System, Symmetry Conditions 58
31. Solution of the Field Equations 60
32. Equations of Geodesics 61
33. Newtonian Motion of a Planet 62
34. Relativity Motion of a Planet 63
35. Deflection of Light 65
36. Shift of Spectral Lines 66

INTRODUCTION

Since we are going to deal with applied Mathematics, or Mathematics applied to Physics we have to state in the beginning the general point of view we take on that subject. A mathematical theory consists of statements or propositions, some of which are written as formulas. Some of these propositions are proved, that means deduced from others, and some are not; the latter are called definitions and axioms. Furthermore most mathematical theories, in particular those in which we are interested, deal with quantities, so that the propositions take the form of relations between quantities.

Physical experiments also deal with quantities which are measured in definite prescribed ways, and then empirical relations are established between these measured quantities.

In an application of Mathematics to Physics a correspondence is established between some mathematical quantities and some physical quantities in such a way that the same relationship exists (as a result of the mathematical theory) between mathematical quantities as the experimentally established relation between the corresponding physical quantities. This view is not new, it was emphatically formulated by H. Hertz in the introduction to his Mechanics, and then emphasized again by A. S. Eddington in application to Relativity. The process of establishing the correspondence between the physical and the mathematical quantities we shall, following Eddington, call identification. An identification is successful, if the condition mentioned above is fulfilled, viz., if the relations deduced for the mathematical quantities are experimentally proved to exist between the Physical quantities with which they have been identified. From this point of view we do not speak of true or false theories, still less of absolute truth, etc.; truth for us is nothing but a successful identification, and it is necessary to say expressly that there may exist at the same time two successful identifications, two theories, each of which may be applied within experimental errors to the known experimental results; and that there may be times when no such theory has been found; and also that an identification which is successful at one time may cease to be so later, when the experimental precision will be increased.

Very often it happens that the quantities of a theory are compared not with quantities which are direct results of experiment but with quantities of another, less comprehensive, theory whose identification with experimental quantities has proved successful. In fact, it seldom happens that we have to deal with direct results of experiment, since even an experimental paper usually contains a great amount of theory.

We may consider Geometry as a first attempt at a study of the outside world. It may be considered as a deductive system which reflects (in the sense explained above, that is of the existence of a correspondence, etc.) very well our experiences with some features of the outside world, namely features connected with the displacements of what we call rigid bodies. We see at once how much is left out in such a study; in the first place, time is almost entirely left out: in trying to bring into coincidence two triangles we are not interested in whether we move one slowly or rapidly; in describing a circle we are not concerned with uniformity of motion. Another important feature that is left out is the distinction between vertical and horizontal lines although we know that this distinction is a very real one. Then optical and electromagnetic phenomena are left out. We see thus that Geometry is not a complete theory of the outside world; one method of building a complete theory would be to introduce corrections into geometry, to introduce one by one time, gravitation, ether, electricity, to patch it up every time we discover a hole; this is a disrespectful but roughly correct description of the actual development of Mathematical Physics. Another method would be to scrap Geometry, and to build instead a new theory which would be an organic whole, embracing the displacement phenomena as a (very important, to be sure) special case. The purpose of the following discussion is to exhibit such a theory. Before we come to the systematic exposition we want to say something about the plan we are going to follow. It is possible to start with a complete statement of the general theory, and then show how special features may be obtained by specializations and approximations; instead, on didactical grounds, we shall begin with special cases and work up by modifications (these being counterparts of approximations) and generalizations. The essential difference between this procedure and the development of classical Physics is that in the latter Geometry was considered as a fixed basis, not to be affected by the upper structure, and we shall feel free to modify geometry when necessary.

Chapter I.

OLD PHYSICS

The purpose of this chapter is to reformulate some of the fundamental equations of mechanics and electrodynamics, and to write them in a new form which constitutes an appropriate basis for the discussion that follows. The contents of the chapter is classical, the modifications which are characteristic of the Relativity theory have not been introduced, but, as was mentioned the form is decidedly new.

1. Motion of a Particle.

The Inverse Square Law.

The fundamental equations of Mechanics of a particle are usually written in the form

1.1 $m \cdot d^2x/dt^2 = X,\ m \cdot d^2y/dt^2 = Y,\ m \cdot d^2z/dt^2 = Z.$

Here m denotes the mass of the particle, x,y,z are functions of the time t whose values are the coordinates of the particles at the corresponding time, and X,Y,Z are functions of the coordinates whose values are the components of the force at the corresponding point. This system of equations was the first example of what we may call mathematical physics, and much that is now mathematical physics may be conveniently considered as a result of a development whose germ is the system 1.1. This chapter will be devoted to tracing out some lines of this development.

We begin by writing the equations 1.1 in the form

1.11 $dmu/dt = X,\ dmv/dt = Y,\ dmw/dt = Z,$

where

1.2 $u = dx/dt,\quad v = dy/dt,\quad w = dz/dt$

are the velocity components. The quantities

$$mu,\qquad mv,\qquad mw$$

are called the momentum components, and in this form our fundamental equations express the statement that the time rate of change of the momentum is equal to the force, the original statement of Newton. Equations 1.11 are seen to be equivalent to 1.1 if we use the notations 1.2 and the fact, usually tacitly assumed, that the mass of a particle does not change with time or in symbols

1.12 $dm/dt = 0.$

In the equations 1.1, x,y,z are usually unknown functions of the time and X,Y,Z are given functions of the coordinates. The situation is then this: first the field has to be described by giving the forces X,Y,Z, and then the motion in the given field is determined by solving equations 1.1 (with some additional initial conditions).

We shall first discuss fields of a certain simple type. One of the simplest fields of force is the so-called inverse square field. The field has a center, which is a singularity of the field; in it the field is not determined; in every other point of the field the force is directed toward the center, (or away from it) and the magnitude of the force is inversely proportional to the distance from the center. As the most common realization of such a field of force we may consider the gravitational field of a mass particle or of a sphere. If cartesian coordinates with the origin at the center are introduced the force components are

1.3 $X = cx/r^3,\quad Y = cy/r^3,\quad Z = cz/r^3,$

where c is a coefficient of proportionality, negative, when we have attraction and positive in the case of repulsion, and

1.4 $r^2 = x^2 + y^2 + z^2;$

taking the sum of squares of X,Y,Z we easily find c^2/r^4 so that the magnitude of the force is c/r^2; force is inversely proportional to the square of the distance. If the field is produced by several attracting particles the force at every point (outside of the points where the particles are located) is considered to be given by the sum of the forces due to the separate particles. In this case the expressions become quite complicated and it is easier to study the general properties of such fields by using certain differential equations to which the force-components are subjected rather than by studying the explicit expressions.

These differential equations are as follows:

1.51 $\partial X/\partial x + \partial Y/\partial y + \partial Z/\partial z = 0$

1.52 $\partial Y/\partial z - \partial Z/\partial y = 0,\quad \partial Z/\partial x - \partial X/\partial z = 0,$

$$\partial X/\partial y - \partial Y/\partial x = 0.$$

The fact that the functions X,Y,Z given by the formulas 1.3 satisfy these equations may be proved by direct substitution; to facilitate

calculation we may notice that differentiation of 1.4 gives

1.6 $r\cdot\partial r/\partial x = x, \quad r\cdot\partial r/\partial y = y, \quad r\cdot\partial r/\partial z = z.$

Differentiating the first of 1.3, we have now

$$\partial X/\partial x = c/r^3 - 3cx/r^4\cdot\partial r/\partial x = c/r^3 - 3cx^2/r^5.$$

Substituting this and two analogous expressions into 1.51 we easily verify it. The verification of 1.52 hardly presents any difficulty.

It is known that equations 1.52 give a necessary and sufficient condition for the existence of a function φ of which X,Y,Z are partial derivatives. The derivative $\partial X/\partial x$ is then the second derivative of this function, and the system 1.51-1.52 may be replaced by the equivalent system

1.53 $X = \partial\varphi/\partial x, \quad Y = \partial\varphi/\partial y, \quad Z = \partial\varphi/\partial z$

1.54 $\partial^2\varphi/\partial x^2 + \partial^2\varphi/\partial y^2 + \partial^2\varphi/\partial z^2 = 0.$

The last equation is known as the equation of Laplace.

In the particular case where X,Y,Z are given by the formulas 1.3 a function φ of which X, Y,Z are partial derivatives is (as it is easy to verify)

$$\varphi = -c/r.$$

We may say now that the field of force given by 1.3 satisfies the differential equations 1.51, 1.52 or 1.53, 1.54 which express the same thing. It is easy to show that these equations are satisfied not only by the field produced by one particle at the origin, but also by that due to any number of particles: first we notice that if a particle is not at the origin this results only in additive constants in the coordinates, and so does not affect partial derivatives which appear in the equations 1.51-1.52 which therefore remain true in this case; secondly, these equations are linear and homogeneous, as a consequence of which the sum of two solutions of these equations necessarily is a solution, or if two fields satisfy these equations their sum also satisfies them; if then, as is generally assumed, the field produced by several particles is the sum of the fields due to the individual particles, such a field also satisfies equations 1.51, 1.52.

Conversely, it can be proved that any field satisfying the differential equations 1.51-1.52 may be produced by a - finite or infinite - set of particles each of which acts according to the inverse square law. We shall not prove here this fact (the proof is given in Potential Theory), but we shall show that these equations furnish us back the inverse square law, if we add the condition that the field must be symmetric with respect to one point.

A situation of this character, a situation where we have to solve a system of partial differential equations with the "additional condition" of symmetry, will appear again later, and in order to be clear about its significance then, it is desirable to treat here this special case in detail.

To begin with we take the system 1.51, 1.52 in the form 1.53, 1.54, i.e., we state that X, Y,Z are derivatives of a function φ. Then from the condition of symmetry of the X,Y,Z field it follows that the field represented by the function φ also must be symmetric, i.e., that this function may depend only on the distance from the origin, because two points which are equidistant from the origin are symmetric with respect to it and, therefore, φ must have the same value in two such points.

We have thus to solve equation 1.54 with the additional condition that φ depends on x, y,z, only through r. Indicating differentiation with respect to r by ' we have

1.56 $\partial\varphi/\partial x = \varphi'\cdot\partial r/\partial x$, etc.

and

1.57 $\partial^2\varphi/\partial x^2 = \varphi''\cdot(\partial r/\partial x)^2 + \varphi'\cdot\partial^2 r/\partial x^2$, etc.

Squaring each of the formulas 1.6 and taking the sum we have

$$r^2\left[\left(\frac{\partial r}{\partial x}\right)^2 + \left(\frac{\partial r}{\partial y}\right)^2 + \left(\frac{\partial r}{\partial z}\right)^2\right] = x^2 + y^2 + z^2$$

or

$$\left(\frac{\partial r}{\partial x}\right)^2 + \left(\frac{\partial r}{\partial y}\right)^2 + \left(\frac{\partial r}{\partial z}\right)^2 = 1;$$

on the other hand differentiating each formula 1.6 we have

$$\left(\frac{\partial r}{\partial x}\right)^2 + r\cdot\frac{\partial^2 r}{\partial x^2} = 1, \text{ etc.};$$

summing these we obtain

$$r\left(\frac{\partial^2 r}{\partial x^2} + \frac{\partial^2 r}{\partial y^2} + \frac{\partial^2 r}{\partial z^2}\right) = 2.$$

Using now 1.57 we can give the equation 1.54 the form

$$\varphi'' + \frac{2}{r}\varphi' = 0 \quad \text{or} \quad \frac{\varphi''}{\varphi'} + \frac{2}{r} = 0$$

whence $\varphi' = \frac{c}{r^2}$, which substituted in 1.56 together with 1.6 gives 1.3. We see thus that the general equations 1.51-1.52 give us all the general information we need about the fields of force in question; we shall call this system the system of equations of a Newtonian field or simply the Newtonian system, although Newton never considered the differential equations that make it up.

We may comment briefly on the mathematical character of the magnitudes and equations we

have been dealing with in this section. At every point X,Y,Z may be considered as the components of a vector, the force vector; we have thus a vector at every point of space, and this constitutes what is called a vector field. The function φ is an example of a scalar field. These two fields are in the particular relation that the first is derived from the second by differentiation. The vector field satisfies equation 1.51, the left hand side of which is called divergence of the vector field. To find a divergence of a vector field we take the sum of the derivatives of the components of the vector with respect to the corresponding coordinates, or we differentiate each component with respect to the corresponding coordinate and add up the results. The formula becomes more expressive if we number the coordinates by writing

$$1.71 \qquad x = x_1, \quad y = x_2, \quad z = x_3$$

and also the vector components, viz.,

$$1.72 \qquad X = X_1, \quad Y = X_2, \quad Z = X_3.$$

The expression for the divergence may then be written as

$$1.8 \qquad \sum_{i=1}^{3} \partial X_i / \partial x_i.$$

The operation of forming a divergence is of fundamental importance in what follows.

We abandon now for a while the study of force fields and direct our attention to the left hand sides of the equations of motion.

2. Two Pictures of Matter.

Our fundamental equations 1.1 connect matter as represented by the left hand sides with forces as represented by the right hand sides. There seems to be a fundamental difference in the mathematical aspects of matter and force. The quantities characterizing matter, the momentum components, for example, are functions of one variable t, and are subjected to ordinary differential equations, whereas quantities characterizing force, X,Y,Z are functions of three variables x,y,z and, as a consequence, are subjected to partial differential equations; they are field quantities, whereas the matter components are not; another way of saying this is to say that force seems to be distributed continuously through space but matter seems to be connected with discrete points. This distinction is, however, not as essential as it looks; it is merely the result of the point of view we take. We could very well consider matter to be distributed continuously through space; each of the two theories, the discrete theory, according to which matter consists of discrete particles, or material points, each of which carries a finite mass, and the continuous theory, according to which matter is distributed continuously through space, or certain portions of space, may be considered as the limiting case of the other. We may start with material points, then increase their number at the same time decreasing the mass of each and so approximate with any degree of precision a given continuous distribution; or we may start with a continuous distribution, then make the density decrease everywhere except in the constantly decreasing neighborhoods of a discrete number of points, and thus approximate, with any precision a given discrete distribution. It is clear that there cannot be any question as to which of the two theories is correct, since the difference between the two can be made as small as we please, and therefore the predictions based on the two theories can be made to agree as closely as we may wish, so that if one identification is successful within experimental error the other will be likewise. Mathematically the difference will be largely that between ordinary differential equations, which are used in treating the motion of discrete particles, and partial differential equations, which apply to continuous distributions.

We may remark here that although forces are usually considered to be continuously distributed in space, it is possible to introduce a discrete picture here also; this is being actually done sometimes in the electromagnetic theory, when a field of force is represented by discrete lines of force, and the intensity is characterized by the number of lines per square inch; we shall not, however, have occasion to use this picture.

We may also remark here that in the last few years still a third point of view has appeared (in Quantum Theory) which in a way occupies an intermediate position; mathematically the treatment is that used in the continuous case (partial differential equations), but the interpretation is given in terms of discrete particles, the continuous quantities being considered as probabilities of a particle being within a certain volume, and the like. This point of view also will not be used in what follows, and is mentioned here only for the sake of completeness.

We want now to translate the equations of motion 1.11 into the language of the continuous theory. Each point of space (or of a certain portion of space) will be considered as occupied at each moment (or each moment during a certain period) by a material particle. Here we also denote by u,v,w the velocity components of a particle of matter, but here they are also considered as functions of coordinates as well as of time; by

$$u(x,y,z,t), \qquad v(x,y,z,t), \qquad w(x,y,z,t)$$

we understand the velocity components of a particle which at the time t occupies the position x,y,z. The fundamental quantity in this theory corresponding to mass of the discrete theory is density. A particle does not possess any finite mass, a mass corresponds only to a finite volume (at a given time). To a point (at a given time) we assign a density which may be explained as the limit of the mass of a sphere with the center at the given point divided by the volume of that sphere as the radius of the sphere tends toward zero. A better way of putting it is to say that we consider a point function $\rho(x,y,z,t)$ called the density and that the mass of matter occupying a given volume at a given time is the integral

$$\int \rho(x,y,z,t)\cdot dxdydz$$

extended over the given volume. This integral will, in general, be a function of time; the mass in a given volume changes with time because new matter may be coming in and old matter going out, and they do not exactly balance each other. But if we consider a certain volume at a given time, and then consider at other moments the volume which is occupied by the same matter, then the mass of matter in that new volume must be the same. That means that if we consider x, y, z, as functions of t, namely, as the coordinates of the same particle of matter at different times, and if we consider the region of integration as a variable volume but one that is occupied by the same particles of matter at all times, then the integral must be independent of time, or

$$2.1 \qquad \frac{d}{dt}\int \rho(x,y,z,t)\cdot dxdydz = 0.$$

This may be written also in a differential form as

$$2.2 \qquad \partial\rho u/\partial x + \partial\rho v/\partial y + \partial\rho w/\partial z + \partial\rho/\partial t = 0.$$

We indicate two proofs for this fact; first an easy but not rigorous proof.

For an infinitesimal volume $V = dx\cdot dy\cdot dz$ we may consider density as the same in all points of the volume, so that mass will be the product $V\cdot\rho$ and the derivative of this product will be

$$dV/dt\cdot\rho + V\cdot d\rho/dt.$$

Then again, considering V as a product $dx\cdot dy\cdot dz$ we find

$$dV/dt$$

$$= ddx/dt\cdot dy\cdot dz + dx\cdot ddy/dt\cdot dz + dx\cdot dy\cdot ddz/dt.$$

Now set $dx = x_2 - x_1$; then

$$ddx/dt = d(x_2 - x_1)/dt = dx_2/dt - dx_1/dt$$

$$= u_2 - u_1 = \partial u/\partial x\cdot dx;$$

substituting this in the preceding relation we find

$$dV/dt = V(\partial u/\partial x + \partial y/\partial y + \partial w/\partial z),$$

and since

$$d\rho/dt$$

$$= \partial\rho/\partial x\cdot dx/dt + \partial\rho/\partial y\cdot dy/dt + \partial\rho/\partial z\cdot dz/dt + \partial\rho/\partial t$$

$$= \partial\rho/\partial x\cdot u + \partial\rho/\partial y\cdot v + \partial\rho/\partial z\cdot w + \partial\rho/\partial t,$$

the expression for the derivative of mass gives

$$dm/dt = V\cdot(\partial\rho u/\partial x + \partial\rho v/\partial y + \partial\rho w/\partial z + \partial\rho/\partial t)$$

so that constancy of mass is expressed by the condition 2.2.

A rigorous proof would be based on expressing the integral for the moment t' which may be written as $\int\rho(x',y',z',t')dx'dy'dz'$ using as variables of integration the coordinates of the corresponding particles at the moment t. The formulas of transformation would be

$$2.3 \qquad x' = f(x,y,z,t'), \quad y' = g(x,y,z,t'),$$

$$z' = h(x,y,z,t')$$

where $f(x,y,z,t')$ is the x coordinate at the moment t' of the particle which at the moment t was at x,y,z, etc. Using the formulas of transformation of a multiple integral we would obtain

$$\int\rho(f,g,h,t').J.dxdydz$$

where J is the Jacobian of the functions 2.3 and the integration is over the volume occupied at the moment t. Setting the derivative of this integral with respect to t' equal to zero, and then making t' = t and noticing that $\partial f/\partial t'$ for t' = t is the velocity component u, etc., we would find the same equation 2.2.

This equation is called the "continuity equation of matter" or the "equation of conservation of matter." The corresponding equation in the discrete theory is the equation (1.12) $dm/dt = 0$ which is not usually included among the fundamental equations of mechanics. The continuity equation may be written in a very simple form if we use the index notations for the coordinates introduced before (1.71), introduce analogous notations for the velocity components, viz.,

$$2.4 \qquad u = u_1, \quad v = u_2, \quad w = u_3,$$

and in addition write

2.5 $$t = x_4$$

and agree, when it is convenient, to write u_4 for unity so that

2.41 $$1 = u_4.$$

With these notations the continuity equation becomes

2.21 $$\sum_{i=1}^{4} \partial \rho u_i / \partial x_i = 0.$$

Noticing the analogy of this equation with equation 1.8 we are tempted to say that the continuity equation expresses the fact that the "divergence" of the "vector" of components ρu_i is zero. This involves, of course, a generalization of the conceptions divergence and vector, because the summation here goes from one to 4 instead of to 3 as in the above formula. This generalization will be of extreme importance in what follows. In the meantime we may notice that the divergence of the vector ρu_i plays the same role in the continuous theory as the time derivative of the number m played in the discrete theory.

We now continue the translation of the equations of the discrete theory into the continuous language. The equations of motion express the fact that the time derivatives of the momentum components are equal to the force components; limiting our consideration to the left hand sides of the equations we have therefore to consider the time derivatives of the momentum components; in the first place the time derivative of mu; without repeating the reasoning which led us to the continuity equation of matter, and noticing that the only change consists in replacing of m by mu we find that the time rate of change of the first component of the momentum vector will be here

2.61 $$\partial \rho uu/\partial x + \partial \rho uv/\partial y + \partial \rho uw/\partial z + \partial \rho u/\partial t$$

and the analogous expressions for the other components will be

2.62 $$\partial \rho vu/\partial x + \partial \rho vv/\partial y + \partial \rho vw/\partial z + \partial \rho v/\partial t$$

2.63 $$\partial \rho wu/\partial x + \partial \rho wv/\partial y + \partial \rho ww/\partial z + \partial \rho w/\partial t.$$

These expression will have to be set equal to the force components (or, rather, components of the force density) in order to obtain the equations of motion. Such equations have been obtained by Euler for the motion of a fluid and are referred to as Euler's hydrodynamic equations; but at present we are not so much interested in the equations of motion as in the mathematical structure of the expressions involving matter components that we have written down. An attentive inspection will help to discover a far-reaching symmetry which again finds its best expression if we use the index notations introduced above (1.71, 2.4, 2.5). We may write, in fact, for the last three expressions

2.6 $$\sum_{i=1}^{4} \partial \rho u_i u_j / \partial x_i \qquad j = 1,2,3$$

and we note furthermore, that if we let j here take the value 4 we obtain the expression appearing on the left hand side of the equation of continuity (2.21). We come thus to the idea of considering the quantities

2.7 $$M_{ij} = \rho u_i u_j$$

and we see that the expression

2.65 $$\sum_{i=1}^{4} \partial M_{ij} / \partial x_i \qquad j = 1,2,3,4$$

plays a very important part in our theory. The first three components, i.e., the expressions obtained for j = 1,2,3, give the time rate of change of the momentum components, and the last one, obtained by setting j = 4, gives the expression whose vanishing expresses conservation of mass. The expressions 2.65 appear as a generalization of what we call a divergence, and we shall call it divergence also, but it is clear that the whole structure of our expressions deserves a closer study to which we shall devote our next section.

3. Vectors, Tensors, Operations.

We shall later treat the fundamental concepts of vector and tensor analysis in a systematic way. At present we shall show how the language of this theory which for ordinary space has been partly introduced in section 1 can be applied to the case of four independent variables and extended so as to furnish a simple way of describing the relations introduced in the preceding section.

A quantity like ρ which depends on the independent variables x,y,z,t, we shall call a <u>scalar</u> field. The four quantities u_1, u_2, u_3, u_4, we shall consider as the components of a <u>vector</u> (or of a vector field; the latter, if we want to emphasize the dependence on the independent variables). The sixteen quantities $\rho u_i u_j$ furnish an example of a <u>tensor</u> (or tensor field). A convenient way to arrange the components of a tensor is in a square array; for instance,

3.1
$$\begin{array}{cccc} \rho u_1u_1 & \rho u_1u_2 & \rho u_1u_3 & \rho u_1u_4 \\ \rho u_2u_1 & \rho u_2u_2 & \rho u_2u_3 & \rho u_2u_4 \\ \rho u_3u_1 & \rho u_3u_2 & \rho u_3u_3 & \rho u_3u_4 \\ \rho u_4u_1 & \rho u_4u_2 & \rho u_4u_3 & \rho u_4u_4 \end{array}$$

We also obtain a tensor if we differentiate each of the components of a vector with respect to each of the variables x,y,z,t, for instance,

3.2
$$\begin{matrix} \partial u_1/\partial x & \partial u_1/\partial y & \partial u_1/\partial z & \partial u_1/\partial t \\ \partial u_2/\partial x & \partial u_2/\partial y & \partial u_2/\partial z & \partial u_2/\partial t \\ \partial u_3/\partial x & \partial u_3/\partial y & \partial u_3/\partial z & \partial u_3/\partial t \\ \partial u_4/\partial x & \partial u_4/\partial y & \partial u_4/\partial z & \partial u_4/\partial t \end{matrix}$$

We want to mention here a very important tensor the array of whose components is

3.4
$$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

its components are usually denoted by δ_{ij} so that δ_{ij} is one if the indices have the same value, and zero, when they are different. The δ_{ij} are often referred to as the Kronecker symbols.

A tensor has been obtained above from a vector by differentiation; the same process can also be applied to a scalar in order to obtain a vector. From the scalar ρ, we would thus obtain a vector whose components are $\frac{\partial\rho}{\partial x}$, $\frac{\partial\rho}{\partial y}$, $\frac{\partial\rho}{\partial z}$, $\frac{\partial\rho}{\partial t}$; this vector is often called the gradient of the scalar ρ. On the other hand the same process may be applied to a tensor. For instance, differentiating each of the sixteen components, $M_{ij} = \rho u_i u_j$ introduced in 2.7 with respect to each of the independent variables x_i we obtain the 16 x 4 numbers (or functions) $\partial M_{ij}/\partial x_k$. We call these numbers the components of a tensor of rank three, and we may call now what we called simply a tensor, a tensor of rank two, a vector - a tensor of rank one, and a scalar - a tensor of rank zero. The operation of differentiation leads then from a tensor (or, better, a tensor field) to a tensor of the next higher rank.

It is also convenient to introduce another operation, the operation of contraction; it can be applied to a tensor of at least rank two and it lowers the rank by two; for a tensor of rank two it consists in forming the sum of all the components whose indices are equal, or, if the components are arranged as explained before, in taking the sum of all the components in the main diagonal.

The operation of taking the divergence of a vector (field) may be stated now to consist of the operation of differentiation followed by the operation of contraction applied to the resulting tensor of rank two.

A tensor of rank three may be contracted, in general, in three different ways; in general, a tensor of higher rank in as many ways as there are pairs of indices. To contract a tensor with respect to two of its indices means to take the sum of those components in which the two selected indices have the same values; for instance, $\partial M_{ij}/\partial x_k$ is a tensor of rank three; its contraction with respect to the first and the third indices is the sum $\sum_{i=1}^{4} \partial M_{ij}/\partial x_i$; the index j is allowed to take all the four values 1,2,3,4 so that we have four sums which are considered as the components of a vector - the divergence of the tensor M_{ij}.

We may finally mention the operation of multiplication which has been applied several times in what precedes. The vector $\rho u_i = q_i$ has been obtained as a result of multiplication of the vector u_i by the scalar ρ. The tensor M_{ij} has been obtained by multiplying the vector q_i by the vector u_i (every component of the first by every component of the second - that is why we have to use different indices - i and j are supposed to take independently of each of the values 1,2,3,4).

The operation of contraction introduced above will be performed very often; it is convenient, therefore, to simplify our notation; this simplification consists in omitting the symbol of summation, and in indicating that summation takes place by using Greek letters for indices with respect to which we sum. Thus we shall write

3.5 $\partial\rho u_\alpha/\partial x_\alpha = 0$ and 3.6 $\partial M_{i\alpha}/\partial x_\alpha$

for 2.21 and 2.65 respectively. The first gives an example of a divergence of a vector, the second of a divergence of a tensor.

The Greek index in the above formulas has no numerical value; any other Greek letter would do just as well; in this respect a Greek index may be compared with the variable of integration in a definite integral. The only case when we have to pay some attention to the particular Greek letters we are using is when two (or more) summations occur in one expression - in such a case different Greek letters have to be used for every summation. If we have to write, for example, $(\Sigma x_i y_i)^2$ using Greek indices we could write it as $(x_\alpha y_\alpha)^2$, but if we want to write out the two factors instead of using the exponent we have to write $x_\alpha y_\alpha x_\beta y_\beta$ because $(x_\alpha y_\alpha)(x_\alpha y_\alpha)$ would have meant $\Sigma(x_i y_i)^2$.

The operation of contraction is used quite often. The formation of the scalar product of two vectors u_i and v_i may be considered as resulting from their multiplication followed by contraction; the multiplication gives the tensor of the second rank $u_i v_j$, and contracting this we get $u_\alpha v_\alpha = u_1 v_1 + u_2 v_2 + u_3 v_3$, which is the scalar product; the scalar product of two vectors could be also called the contracted product of the two vectors. In an analogous fashion we can form a contracted product of two tensors of the second rank. If the tensors are a_{ij} and b_{ij} the contracted product will be $a_{i\alpha} b_{\alpha j}$; it is also a tensor of rank two. It may be interesting to note that the formation of the contracted product of two tensors is essentially the same operation as that of multiplying two determinants corresponding to the arrays representing the tensors; to see that, it will be enough to consider two three row determinants

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \text{ and } \begin{vmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{vmatrix}$$

Their product, according to theorem of multiplication of determinants is

$$\begin{vmatrix} a_{11}b_{11}+a_{12}b_{21}+a_{13}b_{31} & a_{11}b_{12}+a_{12}b_{22}+a_{13}b_{32} & a_{11}b_{13}+a_{12}b_{23}+a_{13}b_{33} \\ a_{21}b_{11}+a_{22}b_{21}+a_{23}b_{31} & a_{21}b_{12}+a_{22}b_{22}+a_{23}b_{32} & a_{21}b_{13}+a_{22}b_{23}+a_{23}b_{33} \\ a_{31}b_{11}+a_{32}b_{21}+a_{33}b_{31} & a_{31}b_{12}+a_{32}b_{22}+a_{33}b_{32} & a_{31}b_{13}+a_{32}b_{23}+a_{33}b_{33} \end{vmatrix}$$

and it is seen that the elements of this determinant are the components of the tensor of rank two which arises from the tensor a_{ij} and b_{ij} by first multiplying them and then contracting with respect to the two inside indices.

We could also speak of the contracted square of a tensor, meaning by this the contracted product of a tensor with itself.

4. Maxwell's Equations.

In section 1 we discussed from a formal point of view the inverse square law and the fields of a more general nature which can be derived from it; and we expressed the laws of these fields in terms of three-dimensional tensor analysis, i.e., we employed only three independent variables; after that we found that matter is best discussed (from the continuous point of view) by using four-dimensional tensor analysis. We have thus a discrepancy: two different mathematical tools are used in the treatment of the two sides of the fundamental equations of mechanics. This discrepancy will be removed in what follows, it will be removed by considering force fields that differ from those derived by composition of inverse square laws, by modifying in a sense this law; but the modifications will be different in the two cases in which the inverse square law has been applied in older physics, the two cases which we are going to mention now.

Originally, the inverse square law was introduced in the time of Newton in application to gravitational forces; we shall discuss in chapter V the gravitational phenomena, and see what modifications - radical in nature, but very slight as far as numerical values are concerned - the inverse square law will suffer. Later, it has been recognized that the inverse square law applies also to the electrostatic and magnetostatic fields produced by one single electric, resp. magnetic particle. Still later a more general law for electromagnetic fields has been introduced by Faraday and Maxwell, which we shall have to study now.

If X,Y,Z denote the components of electric force in the static symmetric case, as we just said, the inverse square law applies, and, as shown in section 1 it follows under assumption of additivity that for a field produced by any number of particles the divergence vanishes,

$$4.1 \qquad \partial X/\partial x + \partial Y/\partial y + \partial Z/\partial z = 0,$$

and the quantities $\partial Y/\partial z - \partial Z/\partial y$, $\partial Z/\partial x - \partial X/\partial z$, $\partial X/\partial y - \partial Y/\partial x$ also vanish; a <u>static</u> magnetic field does not interact with the electric field, but when a changing magnetic field is present the laws of the electric field are modified; viz., the quantities just mentioned are not zero any more but are proportional or, in appropriately chosen units, equal to the time derivatives of the components L,M,N of the magnetic field, so that we have, in addition

$$4.11 \qquad \begin{aligned} \partial Y/\partial z - \partial Z/\partial y &= \partial L/\partial t \\ \partial Z/\partial x - \partial X/\partial z &= \partial M/\partial t \\ \partial X/\partial y - \partial Y/\partial x &= \partial N/\partial t. \end{aligned}$$

In a similar fashion, the divergence of the magnetic force vanishes,

$$4.2 \qquad \partial L/\partial x + \partial M/\partial y + \partial N/\partial z = 0,$$

and the expressions $\partial M/\partial z - \partial N/\partial y$, $\partial N/\partial x - \partial L/\partial z$, $\partial L/\partial y - \partial M/\partial x$ are proportional to the time derivatives of the electric components, the factor of proportionality, however, <u>cannot be reduced to one</u>; by an appropriate choice of units it can be reduced to <u>minus one</u>, and no changing of directions or sense of coordinate axes can permit us to get rid of this minus sign without introducing a minus sign in the preceding equations; this minus sign is of <u>extreme importance</u> in what follows, as we shall have occasion to observe many times; in the meantime we write out the remaining equations

$$4.21 \qquad \begin{aligned} \partial M/\partial z - \partial N/\partial y &= -\partial X/\partial t \\ \partial N/\partial x - \partial L/\partial z &= -\partial Y/\partial t \\ \partial L/\partial y - \partial M/\partial x &= -\partial Z/\partial t. \end{aligned}$$

Several remarks must be made here concerning these equations, which will be referred to as Maxwell's equations.

In the first place, these equations cannot be proved; they have to be regarded as the fundamental equations of a mathematical theory, whose justification lies in the fact that its quantities have been successfully identified with measured quantities of Physics, in the sense that for physical quantities the same relationships have been established experimentally as those deduced for the corresponding theoretical quantities from the fundamental equations. In the second place, the equations as they appear above present a simplified and idealized form of the fundamental equations, namely the fundamental equations for the case of free space, i.e., complete absence of matter.

In the third place the choice of units which made the above simple form possible concerns not only units of electric and magnetic force, but also units of length and time; it was necessary to choose them in such a way that the velocity of light, which in ordinary units is 300,000 kilometers per second becomes <u>one</u>. As a result of this ordinary velocity, those we observe in everyday life are expressed by very small quantities.

For our purposes it is convenient to arrange our equations in the following form, where differentiation with respect to a variable is indicated by a subscript,

$$4.3 \qquad \begin{array}{ll} N_y - M_z - X_t = 0 & Z_y - Y_z + L_t = 0 \\ L_z - N_x - Y_t = 0 & X_z - Z_x + M_t = 0 \\ M_x - L_y - Z_t = 0 & Y_x - X_y + N_t = 0 \\ X_x + Y_y + Z_z = 0 & L_x + M_y + N_z = 0. \end{array}$$

As mentioned before the above equations describe the behavior of electric and magnetic forces in free space, that is in regions where there is no matter, or where we may neglect matter. On the discrete theory of matter these equations still hold everywhere except at points occupied by matter - in this theory matter appears as singularity of the field and some numerical characteristics of matter, such as electric charge, appear as residues corresponding to these singularities. We shall not discuss this point of view, although mathematically it is very interesting. On the continuous theory of matter some terms which represent matter have to be added to the preceding equations. The second set of Maxwell's equations (4.3) remains unaltered, but the first set is modified; the left hand sides do not vanish any more but are proportional to the velocity components of matter; the coefficient of proportionality is electric density which we denote by ε. The equations of Maxwell for space with matter are thus

$$4.31 \qquad \begin{array}{ll} N_y - M_z - X_t = \varepsilon u & Z_y - Y_z + L_t = 0 \\ L_z - N_x - Y_t = \varepsilon v & X_z - Z_x + M_t = 0 \\ M_x - L_y - Z_t = \varepsilon w & Y_x - X_y + N_t = 0 \\ X_x + Y_y + Z_z = \varepsilon & L_x + M_y + N_z = 0. \end{array}$$

We come thus across a new scalar quantity - electric density. However, in most cases this density is proportional to mass density ρ we have considered before, the factor of proportionality being capable of only two numerical values - one negative for negative electricity, and the other positive for positive electricity.

Even these equations are not sufficient for the description of electromagnetic phenomena; they correspond to a certain idealization in which the dielectric constant and magnetic permeability are neglected, but we shall not go beyond this idealization.

In the above equations we have four independent variables x,y,z,t, as in the discussion of matter in section 2, and we may try now to apply to them the same notations which have been introduced in that section and section 3. The main question here is: how to treat the six quantities X,Y,Z,L,M,N? The question was solved by Minkowski in 1907 in the following way: it is clear that a vector has <u>too few</u> components to take care of these quantities; instead of using two vectors, Minkowski proposed to use a <u>tensor of rank two</u>; of course, a tensor has <u>too many</u> components; to be exact, it has, in the general case, 16 components - four in the main diagonal, six above, and six below; we set those in the main diagonal zero, and those under the main diagonal equal with opposite sign to those above the main diagonal symmetric to them; in this way we are left with six essentially different components; the restriction just introduced is expressed in one formula

$$4.4 \qquad F_{ij} + F_{ji} = 0;$$

in fact, the elements in the main diagonal correspond to equal indices; if we set $j = i$ the above formula becomes $F_{ii} + F_{ii} = 2F_{ii} = 0$, whence $F_{ii} = 0$ as asserted. We try now to identify the components of this tensor with our electromagnetic force components in the following way:

$$4.5 \qquad \begin{array}{lll} X = F_{41}, & Y = F_{42}, & Z = F_{43}, \\ L = F_{23}, & M = F_{31}, & N = F_{12}; \end{array}$$

using these notations together with 4.4 according to which, for example, $F_{14} = -X$, we can write the first set 4.3 in the highly satisfactory form,

$$4.6 \qquad \begin{array}{l} \partial F_{12}/\partial x_2 + \partial F_{13}/\partial x_3 + \partial F_{14}/\partial x_4 = 0 \\ \partial F_{23}/\partial x_3 + \partial F_{21}/\partial x_1 + \partial F_{24}/\partial x_4 = 0 \\ \partial F_{31}/\partial x_1 + \partial F_{32}/\partial x_2 + \partial F_{34}/\partial x_4 = 0 \\ \partial F_{41}/\partial x_1 + \partial F_{42}/\partial x_2 + \partial F_{43}/\partial x_3 = 0 \end{array}$$

or $$\partial F_{i\alpha}/\partial x_\alpha = 0.$$

These four equations show a high degree of symmetry; moreover they show a very pronounced similarity to some of the equations we have been considering in section 2, and for which we prepared a mathematical theory in section 3; we can say that the four equations written above express the fact that the divergence of the tensor F_{ij} just introduced vanishes in the case of free space. However, if we apply the same notations to the second set 4.3 of Maxwell's equations nothing very simple comes out; the minus sign mentioned after formula 4.2 above seems to cause trouble; but there exists a way out also from this difficulty; it has been indicated (before Minkowski's paper) in the work of Poincare and Marcolongo, and foreshadowed in a private

letter by Hamilton as early as 1845. We can overcome the difficulty if we allow ourselves to use imaginary quantities side by side with real quantities - this ought to cause no worry provided we know the formal rules of operations, since our new notations are of an entirely formal nature anyway; we set now instead of 2.5

$$4.7 \qquad x = x_1, \quad y = x_2, \quad z = x_3, \quad it = x_4$$

and instead of 4.5

$$4.72 \qquad \begin{aligned} iX &= F_{41}, \quad iY = F_{42}, \quad iZ = F_{43}, \\ L &= F_{23}, \quad M = F_{31}, \quad N = F_{12} \end{aligned}$$

and then the first set (4.3) becomes (4.5) as before, but the second set (4.3) also acquires a highly satisfactory form, namely

$$4.61 \qquad \begin{aligned} \frac{\partial F_{42}}{\partial x_2} + \frac{\partial F_{24}}{\partial x_3} + \frac{\partial F_{32}}{\partial x_4} &= 0 \\ \frac{\partial F_{34}}{\partial x_1} + \frac{\partial F_{41}}{\partial x_3} + \frac{\partial F_{13}}{\partial x_4} &= 0 \\ \frac{\partial F_{42}}{\partial x_1} + \frac{\partial F_{14}}{\partial x_2} + \frac{\partial F_{21}}{\partial x_4} &= 0 \\ \frac{\partial F_{23}}{\partial x_1} + \frac{\partial F_{31}}{\partial x_2} + \frac{\partial F_{12}}{\partial x_3} &= 0 \end{aligned}$$

$$\text{or} \qquad \frac{\partial F_{ij}}{\partial x_k} + \frac{\partial F_{jk}}{\partial x_i} + \frac{\partial F_{ki}}{\partial x_j} = 0.$$

As mentioned before we consider the components F_{ij} as the components of a tensor. We may say that we have sixteen of them; the six which appear in the relations 4.72, six more which result from them by interchanging the indices and whose values differ from those given in 4.72 only in sign, and four more with equal indices. According to the formula (4.4) they are zero. We may arrange them in a square array as follows:

$$4.8 \qquad \begin{array}{cccc|cccc} F_{11} & F_{12} & F_{13} & F_{14} & 0 & N & -M & -iX \\ F_{21} & F_{22} & F_{23} & F_{24} & -N & 0 & L & -iY \\ F_{31} & F_{32} & F_{33} & F_{34} & M & -L & 0 & -iZ \\ F_{41} & F_{42} & F_{43} & F_{44} & iX & iY & iZ & 0 \end{array}$$

We may compare the property $F_{ij} = -F_{ji}$ which our new tensor has with the property $M_{ij} = M_{ji}$ possessed by the tensor of matter (2.7) and which is simply the result of commutativity of multiplication. These two properties are manifested in the square arrays (3.1 and 4.8) in that the components of M_{ij} which are symmetric with respect to the main diagonal are equal, and those of F_{ij} which are symmetric with respect to the main diagonal are opposite. Tensors of the first type are called symmetric, those of the second - antisymmetric.

We want to see now whether the notations which permitted us to write in a nice form Maxwell's equations would not spoil the nice form which we previously gave to the hydrodynamical equations. But since here we have at our disposal the quantities u_1, u_2, u_3, u_4 we can arrange it so as to offset the i in the x_4. In fact, since differentiation with respect to time occurs always in the presence of an u_4 it is enough to set

$$4.9 \qquad u_4 = i$$

instead of 1, as in 2.41, and everything will be all right, as far as the left hand sides of the equations are concerned, except that the left hand side of the continuity equation becomes imaginary. This, however, does not matter since the right hand sides of the equations, we temporarily disregard.

And now we may consider the Maxwell equations with matter 4.31; the second set is not affected and may be written as 4.61, but the first will appear in a form which may be written simply as

$$4.62 \qquad \partial F_{i\alpha}/\partial x_\alpha = \varepsilon u_i.$$

The equation of continuity of matter is a consequence of these equations; to obtain it differentiate, contract and use the property $F_{ij} = -F_{ji}$; the result gives the continuity equation of matter if we take into account that ρ/ε is a constant.

5. The Stress-Energy Tensor.

So far the equations to which we subjected force components have been linear equations, whereas operations performed on matter involved squares and products of matter components; the similarity which we observed in the mathematical aspects of force and matter components makes it seem desirable to subject force components to operations analogous to those which we applied to matter, viz., multiplication.

In the static case we return to our notations (1.72) X_1, X_2, X_3 for X,Y,Z and form the tensor X_iX_j; this tensor, slightly modified plays an important part in the theory; the modification consists in subtracting from it $\frac{1}{2}\delta_{ij}X_\alpha X_\alpha$ where δ_{ij} are the components of the tensor introduced in 3.4 and $X_\alpha X_\alpha$ stands for the contracted square of the vector X_i, i.e., $X^2 + Y^2 + Z^2$. We consider then the tensor

$$X_iX_j - \tfrac{1}{2}\delta_{ij}X_\alpha X_\alpha$$

whose array is

$$\begin{array}{ccc} \frac{1}{2}(X^2 - Y^2 - Z^2) & XY & XZ \\ XY & \frac{1}{2}(Y^2 - X^2 - Z^2) & YZ \\ XZ & YZ & \frac{1}{2}(Z^2 - X^2 - Y^2). \end{array}$$

Of this tensor we form the divergence (three-dimensional), and find as its components

$$X(X_x + Y_y + Z_z) + Y(X_y - Y_x) + Z(X_z - Z_x),$$
$$X(Y_x - X_y) + Y(X_x + Y_y + Z_z) + Z(Y_z - Z_y),$$
$$X(Z_x - X_z) + Y(Z_y - Y_z) + Z(X_x + Y_y + Z_z);$$

the connection of these expressions with the Newtonian equations (1.51, 1.52) is obvious; the expressions in brackets are the left hand sides of the Newtonian equation, so that the divergence of our new tensor $X_iX_j - \delta_{ij}X_\alpha X_\alpha$ vanishes as the result of Newtonian equations. This again confirms us in our opinion that from the mathematical point of view force and matter components are of very similar nature. We have now in mind electric and magnetic forces; if X, Y, Z are the components of the electric force vector the tensor whose array is written out above is called the "electric stress tensor"; an analogous expression in magnetic components is called the "magnetic stress tensor"; the sum of the two, namely

$$\begin{array}{lll} \tfrac{1}{2}(X^2+L^2-Y^2-Z^2-M^2-N^2) & XY+LM & XZ+LN \\ XY+LM & \tfrac{1}{2}(Y^2+M^2-X^2-Z^2-L^2-N^2) & YZ+MN \\ XZ+LN & YZ+MN & \tfrac{1}{2}(Z^2+N^2-X^2-Y^2-L^2-M^2) \end{array}$$

is called the "electromagnetic stress tensor"; it has been introduced by Maxwell and plays some part in electromagnetic theory, for instance, in the discussion of light pressure; but its main applications and importance seem to be in the study of the fundamental questions, as part of a more general four-dimensional tensor.

We saw how nicely the system of indices worked in the case of Maxwell's equations; it is natural to express in index notation this tensor also. We assert that the required expression is given by

$$5.1 \qquad E_{ij} = F_{i\rho}F_{\rho j} - \tfrac{1}{4}\delta_{ij}F_{\sigma\rho}F_{\rho\sigma}$$

where i,j take on values 1,2,3, and the summations indicated by ρ and σ are extended from 1 to 4. In fact,

$$F_{\sigma\rho}F_{\rho\sigma} = F_{12}F_{21} + F_{13}F_{31} + F_{14}F_{41} + F_{21}F_{12} + F_{23}F_{32} + \ldots.$$
$$= 2(-L^2 - M^2 - N^2 + X^2 + Y^2 + Z^2),$$

$$F_{1\rho}F_{\rho 1} = F_{12}F_{21} + F_{13}F_{31} + F_{14}F_{41} = -N^2 - M^2 + X^2,$$

$$E_{11} = F_{1\rho}F_{\rho 1} - \tfrac{1}{4}F_{\sigma\rho}F_{\rho\sigma} = -N^2 - M^2 + X^2 + \tfrac{1}{2}L^2 + \tfrac{1}{2}M^2 + \tfrac{1}{2}N^2$$
$$- \tfrac{1}{2}X^2 - \tfrac{1}{2}Y^2 - \tfrac{1}{2}Z^2 = \tfrac{1}{2}(X^2 + L^2 - Y^2 - Z^2 - M^2 - N^2)$$

$$E_{12} = F_{1\rho}F_{\rho 2} = F_{13}F_{32} + F_{14}F_{42} = LM + XY$$

and similarly for the other components corresponding to different combinations of the indices 1,2,3 for i and j. There seems to be an inconsistency here; the summation indices ρ and σ we let run from 1 to 4, but we consider only the values 1,2,3 for i and j. It is interesting to see what comes out if we let i and j take the value 4. We get four new components, namely,

$$E_{14} = F_{1\rho}F_{\rho 4} = F_{12}F_{24} + F_{13}F_{34} = i(NY - MZ),$$

$$E_{24} = F_{2\rho}F_{\rho 4} = F_{21}F_{14} + F_{23}F_{34} = i(LZ - NX),$$

$$E_{34} = F_{3\rho}F_{\rho 4} = F_{31}F_{14} + F_{32}F_{24} = i(MX - LY),$$

$$E_{44} = F_{4\rho}F_{\rho 4} - \tfrac{1}{2}(-L^2 - M^2 - N^2 + X^2 + Y^2 + Z^2)$$
$$= \tfrac{1}{2}(X^2 + Y^2 + Z^2 + L^2 + M^2 + N^2).$$

These quantities happen also to have physical meaning. The first three constitute (except for the factor i) the components of the so-called Poynting vector, and the last one is the so-called electromagnetic energy (or, energy density). We are thus led by the notations we have introduced in a purely mathematical way to some physical quantities; we may say that the entire tensor with its sixteen components - it is called "the electromagnetic stress-energy tensor" - unifies in a single expression all the second degree quantities appearing in the electromagnetic theory; the stress components, the Poynting components, and the energy.

The stress-energy tensor may be written out in the form of the following square array:

$$5.2 \qquad \begin{array}{llll} X^2 + L^2 - h & XY + LM & XZ + LN & i(NY - MZ) \\ XY + LM & Y^2 + M^2 - h & YZ + MN & i(LZ - NX) \\ XZ + LN & YZ + MN & Z^2 + N^2 - h & i(MX - LY) \\ i(NY - MZ) & i(LZ - NX) & i(MX - LY) & h \end{array}$$

5.3 where $h = \tfrac{1}{2}(X^2 + Y^2 + Z^2 + L^2 + M^2 + N^2)$.

6. General Equations of Motion. The Complete Tensor.

We let ourselves be guided once more by what seems to be natural from the formal point of view, and form the divergence (four-dimensional divergence) of the new tensor. This can be done either in components or in index notation. We show how to do it the latter way, and leave it to the reader to write out the stress-energy tensor as an array and to form the divergence of the separate lines. Applying formula 3.6 to the tensor 5.1 we have

$$\frac{\partial E_{i\gamma}}{\partial x_\gamma} = \frac{\partial F_{i\alpha}}{\partial x_\gamma}F_{\alpha\gamma} + F_{i\alpha}\frac{\partial F_{\alpha\gamma}}{\partial x_\gamma} - \frac{1}{4}\delta_{i\gamma}\left(2F_{\beta\alpha}\frac{\partial F_{\alpha\beta}}{\partial x_\gamma}\right)$$
$$= \frac{\partial F_{i\alpha}}{\partial x_\gamma}F_{\alpha\gamma} + F_{i\alpha}\frac{\partial F_{\alpha\gamma}}{\partial x_\gamma} + \frac{1}{2}F_{\alpha\beta}\frac{\partial F_{\alpha\beta}}{\partial x_i}$$

the first term on the right may be split up in two equal parts, one of which, writing β for γ, may be written as $\frac{1}{2}\frac{\partial F_{i\alpha}}{\partial x_\beta}F_{\alpha\beta}$ and the other,

writing α for γ and β for α, and interchanging indices in both factors, which does not affect the value because it amounts to changing the sign twice, takes the form $\frac{1}{2}\frac{\partial F_{\beta i}}{\partial x_\alpha}F_{\alpha\beta}$. We thus have

$$\frac{\partial E_{i\gamma}}{\partial x_\gamma} = \frac{1}{2}F_{\alpha\beta}\left(\frac{\partial F_{i\alpha}}{\partial x_\beta} + \frac{\partial F_{\alpha\beta}}{\partial x_i} + \frac{\partial F_{\beta i}}{\partial x_\alpha}\right) + F_{i\alpha}\frac{\partial F_{\alpha\gamma}}{\partial x_\gamma}.$$

Substituting for the second factors their values from Maxwell's equations 4.61 and 4.62 (in space with matter) we get

$$\frac{\partial E_{i\gamma}}{\partial x_\gamma} = \varepsilon F_{i\alpha}u_\alpha$$

or in components without indices

$$\begin{aligned}
\frac{\partial E_{1\gamma}}{\partial x_\gamma} &= \varepsilon(Nv - Mw + X),\\
\frac{\partial E_{2\gamma}}{\partial x_\gamma} &= \varepsilon(Lw - Nu + Y),\\
\frac{\partial E_{3\gamma}}{\partial x_\gamma} &= \varepsilon(Mu - Lv + Z),\\
\frac{\partial E_{4\gamma}}{\partial x_\gamma} &= \varepsilon i(Xu + Yv + Zw).
\end{aligned} \qquad 6.1$$

These expressions obtained by us in a purely formal way are known to possess physical significance: the first three represent the components of the force exerted by an electromagnetic field on a body of electric charge density ε, and the last (if we neglect the factor i) is the rate at which energy is expended by the field in moving the body. The first three expressions obviously give the right hand sides of our hydrodynamic equations. Since the left hand sides also (2.61, 2.62) have been obtained as divergence components (3.6) we may write these equations in an extremely simple form, if we introduce a new tensor, the difference of the two appearing on the left and right hand sides, viz.,

$$T_{ij} = M_{ij} - E_{ij}. \qquad 6.2$$

Our hydrodynamical equations are then simply

$$\partial T_{i\alpha}/\partial x_\alpha = 0 \qquad (i = 1,2,3). \qquad 6.3$$

This seems to be very satisfactory, but there is an unpleasant feature about it, namely, that for i = 4 we do not seem to get a correct equation: the contribution of the tensor M_{ij} gives the left hand side of the continuity equation, and is therefore, zero; but the contribution of the stress-energy tensor is the work performed by the forces on the particle and is, in general, not zero. The source of this unpleasantness and the way to remove it will be clear after the reader becomes acquainted with the contents of the next chapter.

It is time now to cast a glance on the situation as it has been worked out till now. We have ten fundamental quantities, ρ,u,v,w,X,Y,Z, L,M,N. They satisfy certain equations, the Maxwell equations (4.61, 4.62), the equation of continuity (3.5), the equations of motion (6.3); in the last named equations our ten quantities enter in certain combinations which are the components of the tensor T_{ij}. This tensor appears then as a very fundamental one. It may be asked whether it determines the ten quantities which enter into it; if it does, all the quantities we have been considering are, in a general sense, components of one entity, the tensor T_{ij}, and all the equations we have introduced express properties of this tensor - that part of Physics which we are discussing in this book, with the exception of the gravitational field, appears then as the study of the tensor T_{ij}. It can be proved that T_{ij} with certain restrictions determines the quantities ρ,u,v,w,X,Y,Z,L,M,N, and in Chapter V it will be shown that the gravitational phenomena also are taken care of by it.

Chapter II.

NEW GEOMETRY

In the preceding chapter we achieved by introducing appropriate notations a great simplicity and uniformity in our formulas. The notations in which indices take the values from 1 to 4 are modeled after those previously introduced in ordinary geometry, the two points of distinction being first that we have four independent variables instead of the three coordinates, and second, that the fourth variable is assigned imaginary values. In spite of these distinctions the analogy with ordinary geometry is very great, and we shall profit very much by pushing this analogy as far as possible, and using geometrical language, as well as notations modeled after those of geometry.

Physics seems to require then, a mathematical theory analogous to geometry and differing from it only in that it must contain four coordinates, one of which is imaginary. The first purpose of this chapter will be to build a theory to these specifications. The remaining part of the chapter will be devoted to a more systematic treatment of tensor analysis.

7. Analytic Geometry of Four Dimensions.

In the present section we shall give a brief outline of properties which we may expect from a four-dimensional geometry guided by analogy with two and three-dimensional geometries; of course, we shall lay stress mainly on those features which we shall need for the application to Physics that we have in mind. In this outline we shall disregard the fact that our fourth coordinate must be imaginary; certain peculiarities connected with this circumstance will be treated in section 10.

The equations of a straight line we expect to be written in the form

$$7.1 \qquad \frac{x_1-a_1}{v_1} = \frac{x_2-a_2}{v_2} = \frac{x_3-a_3}{v_3} = \frac{x_4-a_4}{v_4}$$

entirely similar to that used in solid analytic geometry; but we may also use another form; as written out the equations state that for every point of the line the four ratios have the same value; denoting this value by p we may express the condition that a point belongs to the line by stating that its coordinates may be written as

$$7.11 \qquad \begin{aligned} x_1 &= a_1+pv_1, \quad & x_2 &= a_2+pv_2, \\ x_3 &= a_3+pv_3, \quad & x_4 &= a_4+pv_4; \end{aligned}$$

giving here to p different values we obtain (for given a_i, v_i) the coordinates of all the different points of the line. The variable p is called the parameter, and this whole way of describing a line is called "parametric representation". Parametric representation is by no means peculiar to four dimensions, it may be, and is, often used in plane and solid analytic geometry. We present it here because we shall need it later, and it is not always sufficiently emphasized.

A straight line is determined by two points; the equations of the line through the points a_i and b_i is given by the above equations (7.1) in which

$$7.2 \qquad v_1 = b_1 - a_1, \quad v_2 = b_2 - a_2, \quad \text{etc.}$$

Two points determine a directed segment or vector, whose components are the differences between the corresponding coordinates of the points, so that v_i is the component of the vector whose initial point is given by a_i and whose final point is given by b_i.

A vector determined by two points of a straight line is said to belong to that line, and we may say that we can use as denominators in the equations 7.1, or as coefficients of p in 7.11 the components of any vector belonging to the straight line.

Two vectors are considered equal if they have equal components; a vector is multiplied by a number by multiplying its components by that number, and two vectors are added by adding their corresponding components.

Two lines are parallel if they contain equal vectors, and it is easy to see that a condition for parallelism of line 7.11 and

$$7.3 \qquad \frac{X_1-A_1}{V_1} = \frac{X_2-A_2}{V_2} = \frac{X_3-A_3}{V_3} = \frac{X_4-A_4}{V_4}$$

is

$$v_1/V_1 = v_2/V_2 = v_3/V_3 = v_4/V_4 \text{ or } V_i = \alpha v_i$$

where α is a number so that proportionality of components of two vectors means parallelism.

A condition for <u>perpendicularity</u> of these two lines we expect to be the vanishing of the expression

$$7.4 \qquad v_1V_1 + v_2V_2 + v_3V_3 + v_4V_4$$

which is called the <u>scalar product</u> of the two vectors v_i and V_i.

The distance between the points a_1, a_2, a_3, a_4 and b_1, b_2, b_3, b_4 is given by the square root of the expression

7.41 $$v_1^2 + v_2^2 + v_3^2 + v_4^2$$

$$= (a_1-b_1)^2 + (a_2-b_2)^2 + (a_3-b_3)^2 + (a_4-b_4)^2;$$

this distance is also considered as the length of the vector joining a_i and b_i. The expression for the square of the length of a vector may be considered as a special case of the expression 7.4. We may say then, that the square of the length of a vector is the square of the vector, i.e., the product of the vector with itself.

We shall often use Roman letters to denote vectors. The scalar product of the vectors x and y will be denoted by x.y and the square of the vector x by x^2.

A vector whose length, or whose square, is unity we shall call a unit vector. Its components, we would expect, may be considered as the direction cosines of the line on which the vector lies. We also expect that the scalar product of two vectors is equal to the product of their lengths times the cosine of the angle between them; but, of course, an angle between two vectors in four-dimensional space has not been defined, so that we could simply define the angle between two vectors by this property, or by the formula

7.5 $$\cos\varphi = \frac{x.y}{\sqrt{x^2}\sqrt{y^2}}.$$

But if we want the angle to be a real quantity the absolute value of this expression cannot exceed unity, or

$$(x.y)^2 \leq x^2.y^2 \quad \text{or} \quad (x.y)^2 - x^2y^2 \leq 0.$$

If we form the vector $\lambda x + \mu y$ where λ and μ are two numbers, the square of that vector would be

$$\lambda^2x^2 + 2\lambda\mu x.y + \mu^2y^2,$$

and the above inequality, which expresses the fact that the discriminant of this expression is negative is seen to be a consequence of the assumption that a square of a vector is never negative.

A plane we would expect to be determined by three points not in a line, or by two vectors with the same initial point or by two lines through a point. Instead of characterizing a plane by equations we prefer to give it in parametric form; limiting ourselves to a plane through the origin we have

7.12
$$x_1 = pa_1 + qb_1, \qquad x_2 = pa_2 + qb_2,$$
$$x_3 = pa_3 + qb_3, \qquad x_4 = pa_4 + qb_4,$$

or

$$x_i = pa_i + qb_i$$

where a_i and b_i are the coordinates of two points in the plane or the components of two vectors of the plane whose initial points may be considered as at the origin. We shall write this formula also as

$$x = \alpha a + \beta b$$

where x, a, b stand for vectors whose components are x_i, a_i, b_i and where we use Greek letters for parameters in order to avoid confusion with vectors which we denote now by Roman letters.

We always can choose two mutually perpendicular unit vectors as the two vectors determining a plane; if we call these vectors i and j the preceding formula becomes

7.6 $$x = \alpha i + \beta j.$$

The fact that i and j are perpendicular unit vectors may be written as

$$i.j = 0, \quad i^2 = j^2 = 1.$$

It is easy to see that in this case α and β are the projections of x on the directions of i and j, or the scalar products of x with i and j respectively.

Every pair of coordinate axes determines a plane and since six pairs can be formed from four objects we have six coordinate planes.

In the same way that the direction of a straight line is determined by a configuration of two points on it - a vector, the "orientation" of a plane may be determined by the configuration of three points on it - a triangle. A vector is given by its components, which are the lengths of its projections on the coordinate axes; in the same way a triangle may be characterized (to a certain extent) by the areas of its projections on the coordinate planes. If, for example, we take a triangle one of whose vertices is at the origin 0,0,0,0 and the two others at the points x_i and y_i respectively, the areas of the projections will be the six quantities

$$F_{ij} = \frac{1}{2}\begin{vmatrix} x_i & x_j \\ y_i & y_j \end{vmatrix}.$$

It is interesting to compare these numbers, which satisfy the relation

$$F_{ij} + F_{ji} = 0$$

with the components of the tensor F_{ij} (compare 4.4), three of which have been identified with the electric, and the other three with the magnetic force components. Our ten fundamental quantities ρu, ρv, ρw, ρ, X, Y, Z, L, M, N, seem to allow thus a geometrical interpretation; the

first four are considered as the four projections of a part of a straight line on the coordinate axes, the remaining six, as the projections of a part of a plane on the coordinate planes.

A little against our expectations, however, these six quantities F_{ij} are not independent; the reader will easily verify, using the above expressions, that

$$7.7 \qquad F_{12}.F_{34} + F_{31}.F_{24} + F_{23}.F_{14} = 0.$$

We have here a relation that exists in the mathematical theory; at once the question arises: does a corresponding relation hold for the corresponding quantities in the physical theory; according to the formulas 4.5 this would mean

$$L.X + M.Y + N.Z = 0,$$

i.e., perpendicularity of the electric and magnetic force vectors; these vectors are, however, known not to be necessarily perpendicular to each other; our identification is therefore faulty; a slight modification would, however, help to overcome the difficulty; if instead of considering the areas of projections of a triangular contour, we consider the areas of projections of an arbitrary contour, not necessarily a flat one, then the six quantities are independent and the formal analogy holds perfectly.

Returning to the plane we may mention that although it might seem strange at first glance, we should expect that two planes may have only one point in common - a situation which never occurs in three dimensions. An example of two planes with only one common point is given by the x_1x_2 and the x_3x_4 coordinate planes; the common point is, of course, the origin.

Four points not in a plane, or three vectors with a common origin, we expect to determine a "solid" which may be defined as the totality of points of three kinds: (1) points on the lines determined by the given vectors; (2) points on lines joining two points of the first kind; and (3) points of lines joining two points of the second kind. In ordinary geometry a configuration defined in this way exhausts all points, but not so in our four-dimensional geometry; as examples may serve the four coordinate solids, the totalities respectively of points satisfying the relations $x_1 = 0$, $x_2 = 0$, $x_3 = 0$, $x_4 = 0$.

A parametric representation of a solid is analogous to that of a plane. For a solid through the origin we have as such parametric representation

$$7.61 \qquad x = \alpha i + \beta j + \gamma k;$$

where i,j,k have again been chosen as perpendicular unit vectors so that

$$i.j = j.k = k.i = 0; \quad i^2 = j^2 = k^2 = 1.$$

For all possible values of α, β, γ we obtain all the points of the solid through the origin determined by the vectors i,j,k.

Next we consider a configuration determined by five points not in a solid; we obtain all the points of our four-dimensional space. Incidentally, as a generalization of the formulas 7.6 and 7.61 we may write now

$$7.62 \qquad x = \alpha i + \beta j + \gamma k + \delta l;$$

this formula gives the expression of every vector with initial point at the origin in terms of four mutually perpendicular unit vectors. We have, of course,

$$7.8 \qquad i.j = i.k = i.l = j.k = j.l = k.l = 0, \quad i^2 = j^2 = k^2 = l^2.$$

8. Axioms of Four-Dimensional Geometry.

Until now we have been listing some propositions that we may expect to have in four-dimensional geometry. But what <u>is</u> four-dimensional geometry? As an abstract mathematical theory it is just a collection of statements of which some may be taken without proof and considered as axioms and definitions, and the others are deduced from them as theorems. It is not difficult then to pass from our expectations of a four-dimensional geometry to a realization of such a geometry; all we would have to do would be to pick out certain of the propositions listed above and consider them as axioms and to show that the others can be deduced from them.

But in so doing we do not want to include among our axioms propositions involving coordinates. In two and three-dimensional geometry we are accustomed to see analytic geometry based on the study of elementary geometry, that precedes it. We have an idea of what a straight line is before we come to coordinate axes, and we choose three of these pre-existing straight lines to play the part of coordinate axes. We may choose these axes to a certain extent arbitrarily (this arbitrariness being restricted only by our desire to have a rectangular system); the coordinate axes play only an auxiliary role, and it would be awkward to include any reference to a particular coordinate system in the axioms. We look, therefore, among the propositions mentioned in the preceding section for some that are independent of a coordinate system and from which we can reconstruct the whole system.

As our fundamental undefined conception we choose "vector". The axioms that follow will in the main be rules of operations on vectors.

Axiom I. Every two vectors have a sum. Addition is commutative $a + b = b + a$, and associative, i.e., $a + (b + c) = (a + b) + c$; subtraction is unique, i.e., for every two vectors a and b there exists one and only one vector x such that $a = b + x$. It follows that there exists a vector 0 that satisfies the relation $a + 0 = a$ for every a.

Axiom II. Given a number α and a vector a there exists a vector αa or $a\alpha$ which is called their product. The associative law holds in the sense that $\alpha(\beta a) = (\alpha\beta)a$ and also the distributive laws $(\alpha + \beta)a = \alpha a + \beta a$ and $\alpha(a + b) = \alpha a + \alpha b$.

Axiom III. To every two vectors a and b corresponds a number a.b or ab called their (scalar) product. Scalar multiplication is commutative, a.b = b.a; it obeys together with multiplication of vectors by numbers the associative law, $\alpha(a.b) = (\alpha a.b)$; and, together with addition - the distributive law $a.(b + c) = a.b + a.c$.

Before we formulate the next axiom we introduce the definition: the vectors a,b,c,.... are called linearly dependent if there exist numbers α, β, γ,.... not all zero, such that

$$\alpha a + \beta b + \gamma c + \ldots = 0;$$

they are called linearly independent when no such numbers exist.

Axiom IV. There are four independent vectors, but there are no five independent vectors.

Axiom V. If a vector is not zero its square is positive.

To these axioms on vectors we have to add some statements concerning points if we want to have a geometry, and as such we may take:

Axiom VI. Every two points A,B have as their difference a vector, h; or in formulas $B - A = h$, $B = A + h$.

Axiom VII. $(A - B) + (B - C) = A - C$.

The body of propositions which may be deduced from these axioms we call four-dimensional geometry.

In order to prove that the whole of geometry can be deduced from the propositions I-VII we would have to actually deduce it. We shall not do it, of course, but we shall indicate how analytic geometry can be arrived at in the following discussion, which is meant to be entirely formal, i.e., during which it is not intended to invoke our intuition but only the properties stated in the axioms.

By length of a vector we mean the positive square root of the product of the vector by itself: $|a| = \sqrt{a^2}$. Two vectors are considered perpendicular, if their scalar product is zero. A vector is called a unit vector if its length is = 1.

Lemma. There exist four mutually perpendicular unit vectors.

Proof. According to Axiom IV there exist four independent vectors a,b,c,d; i.e., such vectors that there are no four numbers α, β, γ, δ, not all zero such that $\alpha a + \beta b + \gamma c + \delta d = 0$. It follows that $a \neq 0$, because if it were zero, the numbers $\alpha = 1$, $\beta = 0$, $\gamma = 0$, $\delta = 0$ would satisfy the above relation. Call a, multiplied by the reciprocal of its length, i, so that $i = a/|a|$. It is easy to see that i is a unit vector, and that i,b,c,d are independent. Consider the vector $b' = b - i(bi)$. It is easy to see that it is perpendicular to i; multiply b' by the reciprocal of its length and call the result j; then, i and j are two perpendicular unit vectors and i,j,c are independent. Call $c' = c - i(ci) - j(cj)$; this vector is perpendicular to both i and j, and if we multiply it by the reciprocal of its length and call the result k, we have in i,j,k, three perpendicular unit vectors; the fourth vector ℓ may be obtained from d in an analogous way.

Lemma II. Given any vector x we have the identity

$$8.1 \qquad x = i(ix) + j(jx) + k(kx) + \ell(\ell x).$$

Proof. Since every five vectors are dependent (Axiom IV), there exist five numbers α, β, γ, δ, ε not all zero such that

$$\alpha i + \beta j + \gamma k + \delta \ell + \varepsilon x = 0;$$

here ε cannot be zero since otherwise i,j,k,ℓ would be dependent and we know that they are not. Dividing by $-\varepsilon$ we have

$$x = (-\tfrac{\alpha}{\varepsilon})i + (-\tfrac{\beta}{\varepsilon})j + (-\tfrac{\gamma}{\varepsilon})k + (-\tfrac{\delta}{\varepsilon})\ell;$$

multiplying by i, the three last terms on the right vanish as the result of perpendicularity of i to j, k and ℓ and the first term becomes $-\frac{\alpha}{\varepsilon}$, whereas the left hand side is ix; in the same way we prove that $-\frac{\beta}{\varepsilon}$, $-\frac{\gamma}{\varepsilon}$, $-\frac{\delta}{\varepsilon}$ have the values jx, kx, ℓx, respectively, and lemma II is proved.

A set of four perpendicular unit vectors we shall call a set of coordinate vectors. We shall call the quantities $x_1 = ix$, $x_2 = jx$, $x_3 = kx$, $x_4 = \ell x$ the components of x with respect to i,j,k,ℓ. A point O together with a set of coordinate vectors we call a coordinate system. Given a coordinate system we can assign to every point X four coordinates in the following way: denote the vector X - O by x (according to Axiom VI); and call the components of the vector x the coordinates of the point X. If now we choose another origin O' and the same set of coordinate vectors, the coordinates of the point X will be the components of X - O'; but since, according to Axiom VII,

$X - O = (X - O') + (O' - O)$, the old coordinates will be equal to the new coordinates plus the old coordinates of the new origin, and thus a connection is established with ordinary analytic geometry.

Formula 8.1 may be compared with the formula 7.62. It may also be written as

$$8.2 \qquad x = x_1 i + x_2 j + x_3 k + x_4 \ell .$$

9. Tensor Analysis.

We want to substitute now for the preliminary definitions of tensor analysis that were suggested by the formal developments in Chapter I, a definition that is more satisfactory. At that time our point of view was simply that we shall consider symbols with two (or more) indices as the tensor components in a way similar to that of using symbols with one index as vector components. But in case of vectors (in ordinary space) we know what vectors are, and we consider the components as a method of representing that known thing. In the case of tensors we seem to have to take representation as the starting point of our study. The situation seems complicated since the fact that there is not one but that there are many different representations of the same vector, depending on the coordinate system we use leads us to think that the same general situation will obtain in the case of tensors, and the question arises naturally: how shall we be able to find out, given two representations of a tensor, whether it is the same tensor that is represented in the two cases; or, given a representation of a tensor in one system of coordinates how to find the representation of the same tensor in a given other system. In order to be able to answer such questions intelligently we want to introduce the idea of the tensor itself, to put it in the foreground and to consider the components as something secondary. In the beginning we shall limit ourselves to the consideration of two dimensions.

We look then for some entity, of which the components will be constituent parts. The first thing that occurs to our minds in connection with tensors of rank two is, of course, a determinant. It is a single number determined by its elements, or components. However, it cannot be used for our purposes because the determinant does not, in turn, determine its components.

Another instance where two index symbols occur in mathematics is the case of quadratic forms. The equation of a central conic may, for example, be written in the form $ax^2 + 2bxy + cy^2 = 1$; or introducing the notations x_1 for x, x_2 for y, a_{11} for a, a_{12} and a_{21} for b, and a_{22} for c, in the form

$$a_{11}x_1x_1 + a_{12}x_1x_2 + a_{21}x_2x_1 + a_{22}x_2x_2 = 1.$$

Let us consider the left hand side of this equation. Here the tensor components a_{ij} are combined (together with the variables x_1, x_2) into one expression; and they can, to a certain extent, be gotten back from that expression. If we set $x_1 = 1$, $x_2 = 0$, for instance, we get a_{11} as the value of the expression; a_{22} can be gotten in a similar way, but it would be difficult to imagine how a_{12} could be obtained; in fact, it is impossible to get a_{12} from this expression, because two expressions which differ in their form but for which $a_{12} + a_{21}$ has the same value would give the same values for all combinations of values of x_1, x_2. A slight generalization will, however, obviate this difficulty. This generalization is suggested by the equation of the tangent to the above conic and can be written as

$$9.1 \qquad \varphi = a_{11}x_1y_1 + a_{12}x_1y_2 + a_{21}x_2y_1 + a_{22}x_2y_2;$$

here, for instance, setting $x_1 = 1$, $x_2 = 0$, $y_1 = 0$, $y_2 = 1$ we get a_{12}. We shall therefore consider the bilinear form above as the tensor.

If we do that we may free ourselves of coordinates easily. The variables x_1, x_2 and y_1, y_2 may be considered as the components of two vectors, and the above expression 9.1 furnishes us then a numerical value every time these two vectors are given; it may be considered as defining a function φ; the arguments of that function are the two vectors and the values are the numbers calculated by substituting the components of these vectors in the expression 9.1. This functional dependence we may consider as the tensor so that if we want to use another coordinate system we shall have the same vectors given by different components x'_1, x'_2 and y'_1, y'_2 and we expect to find another expression of the same type as 9.1 involving these new components, say

$$9.11 \qquad \varphi = a'_{11}x'_1y'_1 + a'_{12}x'_1y'_2 + a'_{21}x'_2y'_1 + a'_{22}x'_2y'_2$$

which would assign the same values to the two vectors. The coefficients will, of course, be different, and these new coefficients we shall consider as the new components of the same tensor in the new coordinate system.

Let us perform the calculation. If we rotate our axes through an angle φ the old coordinates are expressed in the new coordinates by the formulas

$$9.2 \qquad x_1 = x'_1c - x'_2s, \qquad x_2 = x'_1s + x'_2c,$$

where

$$9.3 \qquad c = \cos\varphi, \qquad s = \sin\varphi;$$

the components of the other vector y_1, y_2 will

be expressed by analogous formulas in terms of the new components of the same vector. Substituting these expressions in the above bilinear form (9.1) we get

$$a_{11}(x'_1c-x'_2s)(y'_1c-y'_2s)+a_{12}(x'_1c-x'_2s)(y'_1s+y'_2c)$$

$$+ a_{21}(x'_1s+x'_2c)(y'_1c-y'_2s)+a_{22}(x'_1s+x'_2c)(y'_1s+y'_2c)$$

which may be written as 9.11 if we give to a'_{ij} the values

$$\begin{aligned} a'_{11} &= a_{11}c^2 + a_{12}cs + a_{21}sc + a_{22}s^2 \\ a'_{12} &= -a_{11}cs + a_{12}c^2 - a_{21}s^2 + a_{22}sc \\ a'_{21} &= -a_{11}sc - a_{12}s^2 + a_{21}c^2 + a_{22}cs \\ a'_{22} &= a_{11}s^2 - a_{12}sc - a_{21}cs + a_{22}c^2. \end{aligned} \qquad 9.21$$

These are the new components of the tensor whose old components are the a_{ij}. The equation of the conic section in the new coordinates has the same form as in the old system. We may say that 9.11 expresses the same functional dependence on the two vectors using their new components, as 9.1 does using their old components.

The components of a tensor change, in general, when we pass from one coordinate system to another but there are certain combinations which do not change; for instance, if we add together the first and the last of the four above equalities we obtain, taking into account that

$$c^2 + s^2 = 1 \qquad 9.4$$

the relation

$$a'_{11} + a'_{22} = a_{11} + a_{22}; \qquad 9.5$$

also it is easy to prove that the expression

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \qquad 9.51$$

is not affected by the substitution of the primed components for the unprimed ones.

Expressions of this kind are called invariants.

We have thus achieved our purpose; although we use coordinates in the definition of a tensor the result is independent of the particular coordinate system used. We can go a step farther, however, and dispense with coordinates altogether in the definition of a tensor. Not every functional dependence of a number φ on two vectors we shall call a tensor, the dependence on each vector must be linear (and homogeneous); by this we mean that the expression involves only first powers of the components of each vector, and no products of components of the same vector; our conception of linearity seems thus to involve components and we are not rid of coordinate systems yet. We want therefore to define linearity independently of coordinate representation and we shall see that the following definition, entirely independent of coordinates leads to the same results.

We say, in general, that $\varphi(x)$ depends on its argument x linearly if

$$\varphi(\lambda x + \mu y) = \lambda\varphi(x) + \mu\varphi(y). \qquad 9.6$$

It is easy to see that linearity defined in terms of coordinates as dependence involving only first powers of the components satisfies this condition. We arrive thus at the following definition of a tensor:

A tensor of rank r is a function which assigns to r vector arguments numerical values the dependence of the value on each argument being linear in the sense of 9.6.

We can prove that an expression of a tensor as a bilinear (or multilinear) form may be gotten back from this general definition. In fact, if given, e.g., a tensor of rank two, i.e., with two vector arguments $\varphi(x,y)$ we substitute for x and y their expressions in terms of components and unit vectors (see 7.6)

$$x = ix_1 + jx_2, \qquad y = iy_1 + jy_2,$$

we may write, using the above definition of a tensor and that of linearity (9.6):

$$\varphi(x,y) = \varphi(ix_1 + jx_2, iy_1 + jy_2)$$

$$= \varphi(i,i)x_1y_1 + \varphi(i,j)x_1y_2 + \varphi(j,i)x_2y_1 + \varphi(j,j)x_2y_2.$$

We see that this expression differs from that given by 9.1 as a bilinear form only in that $\varphi(i,i)$, $\varphi(i,j)$, $\varphi(j,i)$, $\varphi(j,j)$ appear instead of a_{11}, a_{12}, a_{21}, and a_{22}. From this point of view the conception of a tensor is entirely independent of a coordinate system and of components. We obtain tensor components when we introduce a set of coordinate vectors; and transformation of coordinates corresponds to replacing of one set of coordinate vectors by a new set.

We pass now to the consideration of operations on tensors. We had three such operations: multiplication, contraction and differentiation.

Multiplication is simple. If we have two tensors, say $f(x,y)$ and $g(z,u,v)$ we obtain a tensor of rank five by multiplying these two together

$$h(x,y,z,u,v) = f(x,y).g(z,u,v);$$

it is easy to see that the components of h are obtained from the components of f and g in the following fashion

$$H_{ijklm} = f_{ij} \cdot g_{klm}.$$

Next comes contraction. We have defined it in indices; i.e., given the components of a tensor of rank two in a certain coordinate system we have a definite rule for obtaining a scalar, viz., $a_{11} + a_{22}$; the question arises: will we obtain the same scalar if we use another system of coordinates, in other words, is the definition independent of the system of coordinates, is it invariant? Yes, this invariance has been proved above by formula 9.5. We have now the right to use the definition of contraction in terms of components, knowing that it has an intrinsic meaning, that the result is independent of the system of coordinates used.

We are in a position now to answer a question that must have arisen in the mind of the reader. In the preceding chapter we agreed to consider a vector as a tensor of rank one. Here with our new definition of tensor a vector and a tensor of rank one seem to be two entirely different things; but we may consider together with every vector a tensor of rank one which has the same components. To find a tensor $f(x)$ which has the same components as the vector v we have to make $f(i) = v_1$, $f(j) = v_2$, and we have

$$f(x) = x_1 f(i) + x_2 f(j) = x_i v_1 + x_2 v_2$$

so that the value of this tensor $f(x)$ is simply the scalar product of the vector to which it corresponds by the vector argument.

Incidentally, this raises a question as to the nature of the scalar product; if we define it as $x_\alpha v_\alpha$ is it invariant? It may be considered as resulting from two tensors of rank one by first multiplying them and then contracting the resulting tensor of rank two.

We also might at this place say a few words about the symbols of Kronecker δ_{ij}. We may try to consider these symbols as the components of a tensor in some coordinate system. The value of the tensor will then be given by $\delta_{\alpha\beta} x_\alpha y_\beta$ and this is easily seen to be $x_\alpha y_\alpha$, the scalar product of the vectors x and y and thus independent of the coordinate system. We may then speak of the tensor δ_{ij} without mentioning the coordinate system because its components are the same in all coordinate systems.

The square of a vector may be defined as the scalar product of a vector with itself, it is the sum of the squares of the components in any coordinate system.

We next take up the operation of differentiation. Let us begin with a scalar field f; f is a function of the coordinates which we do not put in evidence. The coordinates of a point P are the components of the vector OP which joins the origin to the point in question and they depend in the fashion discussed before on the choice of the coordinate vectors.

After choosing a definite coordinate system we may assign to f in every point a vector by agreeing that the components of this vector should be

9.71 $\quad v_1 = \partial f/\partial x_1$, and $v_2 = \partial f/\partial x_2$;

given another coordinate system the relation between x_i and x'_i being given by formulas 9.2, we can form the derivatives

9.72 $\quad \partial f/\partial x'_1$ and $\partial f/\partial x'_2$

and consider them as the components in the new coordinate system of a vector. The question arises whether this will be the same vector as the one introduced above.

In order to settle this question let us see what the components of the vector whose components in the old system were v_i should be in the new coordinate system. According to formulas 9.2 they are

$$v'_1 = v_1 c + v_2 s = \partial f/\partial x_1 . c + \partial f/\partial x_2 . s$$

$$v'_2 = -v_1 s + v_2 c = -\partial f/\partial x_1 . s + \partial f/\partial x_2 . c.$$

On the other hand

$$\partial f/\partial x'_1 = \partial f/\partial x_1 . \partial x_1/\partial x'_1 + \partial f/\partial x_2 . \partial x_2/\partial x'_1,$$

and since

$$\partial x_1/\partial x'_1 = c, \quad \partial x_2/\partial x'_1 = s$$

we find that $v'_1 = \partial f/\partial x'_1$, and, in the same way we find that $v'_2 = \partial f/\partial x'_2$ which shows that the components 9.71 and 9.72 above, are the components in the two coordinate systems considered <u>of one and the same vector</u>. We proved then that the operation of obtaining a vector by taking as its components the derivatives of a scalar with respect to the coordinate axes is independent of the particular system of coordinates used, that means this operation is invariant.

Before passing to differentiation of a tensor of rank higher than zero we may note that the components v_1, v_2 we obtained may be considered as the components of a tensor of rank one; denoting the components of the argument vector by h_1, h_2 we have as the values of this tensor

$$\partial f/\partial x_1 . h_1 + \partial f/\partial x_2 . h_2;$$

this reminds us of a differential and suggests to write dx_1 for h_1 and dx_2 for h_2; we have then the formula

$$df = \partial f/\partial x_1 . dx_1 + \partial f/\partial x_2 . dx_2,$$

which leads to the interpretation of the differential as a tensor of rank one whose components are the derivatives of the given function.

We next consider a tensor of rank one whose components in the old system are f_1, these components being functions of coordinates. Differentiating with respect to x_j we get

$$\partial f_1/\partial x_1, \qquad \partial f_2/\partial x_1$$

$$\partial f_1/\partial x_2, \qquad \partial f_2/\partial x_2;$$

can we consider these as the components of some tensor of rank two? In other words, if we define a tensor by saying that in the old system it has these components will its components in the new system be obtained by differentiating with respect to the new coordinates of the new components f'_1 of the given tensor? A calculation analogous to the one preceding will convince us that this is so.

We have thus introduced an operation which leads from a tensor of rank zero to one of rank one, and from a tensor of rank one to one of rank two, and we could, continuing in the same way pass from any tensor to one of the next higher rank. In introducing this operation we used components of tensors and coordinates of points, but we proved that the result is the same no matter what particular coordinate system we might have used; the operation of differentiation is thus independent of a coordinate system.

After this detailed treatment of tensors and operations on them in plane geometry it will not be difficult to generalize to higher dimensions. We consider first three-dimensional space - solid geometry - and begin with an equation of a central quadric surface. Using notations similar to those introduced at the beginning of this section it may be written as

$$a_{11}x_1x_1 + a_{12}x_1x_2 + a_{13}x_1x_3$$

9.8 $$+ a_{21}x_1x_2 + a_{22}x_2x_2 + a_{23}x_2x_3$$

$$+ a_{31}x_3x_1 + a_{32}x_3x_2 + a_{33}x_3x_3 = 1,$$

or, using our notations for summation with Greek indices introduced in Section 3, as

$$a_{\rho\sigma}x_\rho x_\sigma.$$

For the same reason as before (in the case of a conic) namely, because not all coefficients of such an expression can be obtained as its values, we introduce a slightly more general expression

9.81 $$a_{\rho\sigma}x_\rho y_\sigma$$

as the <u>tensor</u>; giving in it to the variables the values $x_1 = \delta_{2i}$, $y = \delta_{3i}$ (see definition of the symbol δ under 3.4) we obtain, for instance, the coefficient a_{23}.

In coordinateless notation we were independent of the number of dimensions to begin with, so that we may take over the definition of linearity 9.6 and the definition of tensor following it word for word. In order to effectuate the transition from vector notation to coordinate notation we write now, instead of $x = x_1i + x_2j$, $x = x_\rho i_\rho$, and substituting this and an analogous expression for y into $\varphi(x,y)$ we get using 9.6

$$\varphi(x_\rho i_\rho, y_\sigma i_\sigma) = \varphi(i_\rho, i_\sigma)x_\rho y_\sigma.$$

The notation

$$\varphi(i_p, i_q) = a_{pq}$$

brings us back to formula 9.81. There is no difficulty about tensors of higher ranks; quantities with three indices give rise to trilinear forms, e.g.,

$$a_{\rho\sigma\tau}x_\rho y_\sigma z_\tau,$$

those with four indices - to quadrilinear forms, etc. The definitions of multiplication, contraction and differentiation hardly present any difficulty, but we shall devote some time to the question of transformation of coordinates for three and four-dimensions. In solid analytic geometry the question is usually treated by introducing formulas involving all coordinates at the same time, i.e., formulas permitting to pass at once from one system of coordinates to any other with the same origin; these formulas are quite complicated, they involve nine constants which are not independent, but are connected by six relations, and the corresponding thing for four dimensions would be still more unwieldy; we could handle it by introducing index notations, but we prefer another method. We pass from one system of coordinates to another gradually, in steps, each step involving only two of the coordinates - and one constant - the angle through which we rotate in the corresponding plane. Three such steps are enough to pass from any system to any other in three dimensions. For example, we may first perform a rotation in the xy-plane which brings the x-axis into the new xy-plane; then a rotation in the so obtained yz-plane bringing the y-axis into the new xy-plane, and finally we rotate the so obtained x and y axes until they coincide with the new x and y axes.

The advantage of this point of view will be seen from the following proof of the invariance of the operation of contraction in three dimensions. Given a tensor of rank two by its components a_{11}, a_{12}, a_{13}, a_{21}, etc., the result of contraction, according to our definition in Section 3, is

$$a_{\rho\rho} = a_{11} + a_{22} + a_{33}.$$

If we pass to another coordinate system the components will be changed into some components

a'_{ij} and the result of contraction will be

$$a'_{\rho\rho} = a'_{11} + a'_{22} + a'_{33};$$

in order to prove that contraction has an intrinsic meaning, independent of the system of coordinates we have to prove that the last two expressions are equal. If the transformation involves only the x_1 and x_2 coordinates, but does not involve x_3, then a'_{33} which is the coefficient of $x'_3y'_3$ will be a_{33} which is the coefficient of x_3y_3, because $x'_3 = x_3$, $y'_3 = y_3$, and the other coordinates do not depend on x_3 and y_3; the coefficients a'_{11}, a'_{22}, on the other hand will be transformed by the same formulas (9.21) as in the two-dimensional case because x_1, x_2, y_1, y_2 are transformed by the same formulas (9.2) as before. Therefore, formula 9.5 is applicable, and this together with the fact that $a'_{33} = a_{33}$ establishes the invariance of contraction under a transformation of coordinates involving x_1 and x_2 only. But the same reasoning would apply to transformation involving x_2 and x_3 only, or x_1 and x_3 only, and since we have proved that a general transformation of coordinates may be replaced by a succession of transformations involving each only two coordinates we have proved the invariance of the operation of contraction of a tensor of rank two under a general coordinate transformation.

Following the same principle we could prove the invariance of contraction for tensors of any rank and also, using the fact that the invariance of the operation of differentiation has been proved for two dimensions, prove that it has an intrinsic meaning in three dimensions.

We come now to four dimensions. Here it is easy to prove that a general transformation can be effectuated by a succession of <u>six</u> single rotations, i.e., rotations involving only two axes each; in fact, a rotation in the xt-plane will bring the x-axis into the new xyz-solid; a rotation in the yt-plane will bring the y-axis into the new xyz-solid, a rotation in the zt-plane will bring there the z-axis; now the t-axis coincides with the new t-axis and the x,y,z axes are all in the new xyz-solid and can be brought into coincidence with the new x,y,z axes by three more rotations as we saw before.

The reasoning indicated for the three-dimensional case will, therefore, prove the invariance of the fundamental operations of tensor analysis also for four dimensions.

10. Complications Resulting From Imaginary Coordinate.

The only new feature of the new geometry considered in the preceding sections was this that we have four coordinate axes instead of three; but we still have another departure from ordinary geometry (due in the final count to the "minus sign" in the Maxwell equations), viz., the fact that the fourth coordinate is imaginary. The introduction of imaginaries helped us to obtain a symmetrical form of Maxwell's equations, and seems to be beneficial from this point of view. The formal part of the theory runs now smoothly; but there is a disadvantage in this smoothness, it conceals very important peculiarities, and the present section will be devoted to the consideration of some of these peculiarities.

The discussion may be conveniently attached to the consideration of the expression (comp. 7.41)

$$(x_1-x'_1)^2 + (x_2-x'_2)^2 + (x_3-x'_3)^2 + (x_4-x'_4)^2,$$

which defines the square of the distance between the two points whose coordinates appear in it; or the square of the vector joining these two points.

Since in the above formula the quantities x_4 and x'_4 and, therefore, their difference is imaginary, the fourth square is negative, and the expression may, according to the relative magnitudes of the terms, be positive, negative, or zero. There are thus three types of relative positions of two points, or directions, or of vectors; there are vectors of positive square, those of negative square, and those of zero square. Our geometry is thus more complicated than what we would expect it to be if it would differ from the ordinary geometry in the number of dimensions only. This complication, or this richness of our geometry, far from being an undesirable feature, is, as we shall see, an advantage, because it corresponds to certain features of the outside world, e.g., the existence of both matter and light which are going to be identified with two kinds of vectors. At present we only mention that the momentum vector of a material particle, or the vector of components u,v,w,i are vectors of negative square; in fact, the square of the latter is $u^2+v^2+w^2-1$; the first three terms representing the square of the velocity of the particle which, according to the third remark preceding formula 4.3 is very small compared with unity, the expression is negative.

In some cases it may be desirable to sacrifice the formal advantages accruing from the use of the imaginary coordinate in order to put in evidence the peculiarities we are discussing; it is permissible to go back then to the old notations x,y,z,t, but it becomes necessary to modify the formulas accordingly. (We shall see later how it is possible to use index notations and still avoid imaginaries.) If we denote two points by x,y,z,it and x', y', z', it', instead of by x_i and x'_i the formula for the square of the distance will be

$$10.1 \qquad (x-x')^2 + (y-y')^2 + (z-z')^2 - (t-t')^2,$$

and the formula for the scalar product (see 7.4)

of two vectors given by a,b,c,id, and a',b',c', id',

10.2 $$aa' + bb' + cc' - dd'.$$

We may say that the minus sign appearing in these formulas is the same as the one appearing in the second set of Maxwell's equations (4.21), because it may be traced back to them.

We shall occasionally refer to the quantities x_i and u_i which carry indices and involve the square root of -1 as mathematical coordinates and components, and to the quantities x, y,z,t and a,b,c,d as physical coordinates and components. Four-dimensional space of the character we are studying now, i.e., either characterized by three real and one imaginary coordinates, or with four real coordinates but with scalar multiplication with a minus sign given by formula 10.2 is often called four-dimensional space-time because of the interpretation of the quantities x,y,z and t in ordinary physics.

Without going into detail we may mention a few consequences of the "minus sign". According to our definition, two vectors are considered perpendicular if their scalar product vanishes. But now a scalar product of two equal vectors may vanish, as happens, for instance, in the case of two with components 0,0,1,1. We must say then, that such a vector is perpendicular to itself.

We also may mention that corresponding to the existence of three types of directions, those which correspond to vectors of positive square, negative square and zero square, there are three types of orientations or planes. An orientation may best be characterized by the number of zero directions it contains and it is easy to prove that there are orientations containing two, one or no zero directions.

As the result of existence of vectors of negative square our proof that the cosine of the angle between two vectors as defined by formula 7.5 does not exceed unity in absolute value is not applicable and we would have to consider imaginary angles or else consider the cosine as a hyperbolic cosine, but we shall not go into this question.

The only peculiarity due to the "minus sign" other than the existence of zero square vectors that we shall have to use in what follows is connected with transformation of coordinates.

Formally our transformation formulas remain the same; we may write, for instance,

$$x_3' = x_3 \cos\varphi - x_4 \sin\varphi,$$

$$x_4' = x_3 \sin\varphi + x_4 \cos\varphi;$$

but here x_3 is real and x_4 is imaginary and, of course, we expect the new coordinates to be of the same character so that x_3' will be real and x_4' imaginary. Giving to x_3 and x_4 the values 1 and 0, respectively, we see that cos φ must be real, and sin φ - imaginary. We shall, however, prefer not to use imaginary trigonometric functions; in order to avoid them we introduce a new notation as follows:

10.3 $$\cos\varphi = \sigma, \qquad \sin\varphi = i\tau,$$

where σ and τ are real quantities. The identity $\cos^2\varphi + \sin^2\varphi = 1$ gives for σ and τ the relation

10.4 $$\sigma^2 - \tau^2 = 1.$$

If we prefer to use one number, rather than two numbers connected by a relation, in describing different transformations of coordinates in the x_3x_4 plane we may again resort to trigonometry and interpret σ and τ as the secant and tangent of a real angle ψ:

10.5 $$\sigma = \sec\psi, \qquad \tau = \tan\psi.$$

It must be noted, however, that ψ, so to say, has no geometrical significance.

If in the above formulas of transformation we substitute for x_3 and x_4 their expressions in terms of z and t (4.7) and for cos φ and sin φ their expressions (10.3) in terms of σ and τ we obtain the following formulas of transformation for the physical coordinates:

10.6 $$z' = z\sigma + t\tau, \qquad t' = z\tau + t\sigma.$$

These formulas are called the Lorentz transformation formulas and their physical interpretation will be discussed in the next chapter.

Concluding this section we may mention how our axioms of Section 8 must be modified in order to produce a geometry with the desired peculiarities.

It is clear that our axiom V, according to which a non-zero vector has positive square has to be modified. The proper modification is the following:

Axiom V'. There are orientations containing two zero directions, but there are no orientations containing more than two zero directions.

If we replace axiom V by this axiom and keep the remaining axioms as they were stated in Section 8 we obtain a geometry of the kind desired. In order to show this, we first show the existence of four mutually perpendicular vectors three of which have squares equal to 1, and the fourth one equal to -1. We begin by picking out a plane with two zero square vectors a and b; we assert that $a' = \frac{1}{2}(a + b)$ and $b' = \frac{1}{2}(a - b)$ are two perpendicular vectors with squares of opposite sign; in fact $a'b' = \frac{1}{2}(a^2 - b^2)$ and this is zero because $a^2 = 0$ and $b^2 = 0$; then, $a = a' + b'$, squaring this and keeping in mind that $a'b' = 0$, we have $a^2 = a'^2 + b'^2$; since $a^2 = 0$ it follows that the squares of a' and b' are of opposite

sign. Dividing each of the vectors a' and b' by the square root of the absolute value of its square we obtain two mutually perpendicular vectors whose squares are +1 and -1, which we may call k and ℓ respectively. It is easy now to pick out two more vectors i and j which together with k and ℓ constitute a set of four mutually perpendicular vectors; none of them can have a zero square, because if, say, i^2 should be zero all the vectors of the plane determined by a and i would have zero squares, contrary to axiom V'.

Using these four vectors we can, as in the other case, express every vector in the form (comp. 8.2)

$$10.7 \qquad x = \alpha i + \beta j + \gamma k + \delta \ell .$$

The numbers $\alpha, \beta, \gamma, \delta$ will be considered as the components of this vector. In order to show that we have what we wanted we shall express the square of x in terms of its components. Squaring 10.7 and taking into account that $\ell^2 = -1$ we obtain

$$x^2 = \alpha^2 + \beta^2 + \gamma^2 - \delta^2 .$$

This shows that $\alpha, \beta, \gamma, \delta$ are what we call physical components of the vector; the mathematical components are obtained by setting

$$x_1 = \alpha, \quad x_2 = \beta, \quad x_3 = \gamma, \quad x_4 = i\delta ,$$

and we see that we get the kind of geometry we expect to use in physics.

11. Are the Equations of Physics Invariant?

We return now to physics. In Chapter I we arrived at certain equations that we consider as fundamental; namely, the equation of continuity (3.5)

$$11.1 \qquad \partial \rho u_\alpha / \partial x_\alpha = 0,$$

two sets of Maxwell's equations (4.61 and 4.62)

$$11.2 \qquad \partial F_{ij} / \partial x_k + \partial F_{jk} / \partial x_i + \partial F_{ki} / \partial x_j = 0,$$

$$11.3 \qquad \partial F_{i\alpha} / \partial x_\alpha = \varepsilon u_i ,$$

and the equations of motion (6.3)

$$11.4 \qquad \partial T_{i\alpha} / \partial x_\alpha = 0 \qquad (i = 1,2,3)$$

with (comp. 6.2, 5.1, 2.7)

$$11.5 \qquad T_{ij} = E_{ij} - M_{ij} = F_{i\rho} F_{\rho j} - \tfrac{1}{4} \delta_{ij} F_{\sigma\rho} F_{\rho\sigma} - \rho u_i u_j .$$

The fact that the indices run here from 1 to 4 (except in 11.4) suggested four-dimensional geometry; which we have introduced in Sections 7 and 8; the fact that x_4 in the above equations is imaginary (comp. 4.7) suggested the peculiarities discussed in Section 10. Now that we have followed these suggestions and built a mathematical theory we have to see to what results the application of our new theory leads. In addition to following the suggestions we have introduced into the theory a feature that was not directly called for by physics: we made our theory independent of coordinates. In order to bring out the importance of this fact let us consider for a moment the case of two dimensions and compare plane geometry with a two-way diagram. Both in plane analytic geometry and in a diagram we use coordinate axes, but in geometry the axes play an auxiliary role, we find it convenient to express by referring to axes properties of configurations which exist and can be treated independently of the axes; the same properties can be expressed using any system of coordinate axes. The situation is different in the case when we use a plane as a means of representing a functional dependence between two quantities of different kind, when we have a diagram. We may use, for instance, the two axes to plot temperature and pressure, or the height of an individual and the number of individuals of that height. In the majority of such cases the axes play an essential part in the discussion; if we delete the axes the diagram loses its meaning, the question of rotation of axes does not arise.

Returning to physics we have to ask ourselves what we actually need for it, a diagram or a geometry; in other words, are the coordinate axes essential or can they be changed at will, or again, do the equations of physics express properties independent of the coordinate axes; are they invariant, or not.

In order to answer this question let us first consider the formal structure of the equations 11.1 to 11.5. The fundamental dependent variables are here a scalar ρ, a vector u_i and an anti-symmetric tensor F_{ij}.

The left hand side of equation 11.1 may be described as a result of multiplying the scalar ρ by the vector u_i, then differentiating the resulting vector ρu_i, then contracting the tensor so obtained; since the operations of multiplication, differentiation and contraction have been shown to be invariant, the scalar $\partial \rho u_\alpha / \partial x_\alpha$ is independent of the system of coordinates used, and if it is zero in one system of coordinates it is zero in all system of coordinates. The continuity equation expresses, therefore, a fact independent of the system of coordinates employed.

An analogous reasoning applied to 11.3 would show the invariant character of that system. The question of invariance of the first

sign. Dividing each of the vectors a' and b' by the square root of the absolute value of its square we obtain two mutually perpendicular vectors whose squares are +1 and -1, which we may call k and ℓ respectively. It is easy now to pick out two more vectors i and j which together with k and ℓ constitute a set of four mutually perpendicular vectors; none of them can have a zero square, because if, say, i^2 should be zero all the vectors of the plane determined by a and i would have zero squares, contrary to axiom V'.

Using these four vectors we can, as in the other case, express every vector in the form (comp. 8.2)

$$10.7 \qquad x = \alpha i + \beta j + \gamma k + \delta \ell .$$

The numbers $\alpha, \beta, \gamma, \delta$ will be considered as the components of this vector. In order to show that we have what we wanted we shall express the square of x in terms of its components. Squaring 10.7 and taking into account that $\ell^2 = -1$ we obtain

$$x^2 = \alpha^2 + \beta^2 + \gamma^2 - \delta^2 .$$

This shows that $\alpha, \beta, \gamma, \delta$ are what we call physical components of the vector; the mathematical components are obtained by setting

$$x_1 = \alpha, \quad x_2 = \beta, \quad x_3 = \gamma, \quad x_4 = i\delta ,$$

and we see that we get the kind of geometry we expect to use in physics.

11. Are the Equations of Physics Invariant?

We return now to physics. In Chapter I we arrived at certain equations that we consider as fundamental; namely, the equation of continuity (3.5)

$$11.1 \qquad \partial \rho u_\alpha / \partial x_\alpha = 0 ,$$

two sets of Maxwell's equations (4.61 and 4.62)

$$11.2 \qquad \partial F_{ij} / \partial x_k + \partial F_{jk} / \partial x_i + \partial F_{ki} / \partial x_j = 0 ,$$

$$11.3 \qquad \partial F_{i\alpha} / \partial x_\alpha = \varepsilon u_i ,$$

and the equations of motion (6.3)

$$11.4 \qquad \partial T_{i\alpha} / \partial x_\alpha = 0 \qquad (i = 1,2,3)$$

with (comp. 6.2, 5.1, 2.7)

$$11.5 \qquad T_{ij} = E_{ij} - M_{ij} = F_{i\rho} F_{\rho j} - \tfrac{1}{4} \delta_{ij} F_{\sigma\rho} F_{\rho\sigma} - \rho u_i u_j .$$

The fact that the indices run here from 1 to 4 (except in 11.4) suggested four-dimensional geometry; which we have introduced in Sections 7 and 8; the fact that x_4 in the above equations is imaginary (comp. 4.7) suggested the peculiarities discussed in Section 10. Now that we have followed these suggestions and built a mathematical theory we have to see to what results the application of our new theory leads. In addition to following the suggestions we have introduced into the theory a feature that was not directly called for by physics: we made our theory independent of coordinates. In order to bring out the importance of this fact let us consider for a moment the case of two dimensions and compare plane geometry with a two-way diagram. Both in plane analytic geometry and in a diagram we use coordinate axes, but in geometry the axes play an auxiliary role, we find it convenient to express by referring to axes properties of configurations which exist and can be treated independently of the axes; the same properties can be expressed using any system of coordinate axes. The situation is different in the case when we use a plane as a means of representing a functional dependence between two quantities of different kind, when we have a diagram. We may use, for instance, the two axes to plot temperature and pressure, or the height of an individual and the number of individuals of that height. In the majority of such cases the axes play an essential part in the discussion; if we delete the axes the diagram loses its meaning, the question of rotation of axes does not arise.

Returning to physics we have to ask ourselves what we actually need for it, a diagram or a geometry; in other words, are the coordinate axes essential or can they be changed at will, or again, do the equations of physics express properties independent of the coordinate axes; are they invariant, or not.

In order to answer this question let us first consider the formal structure of the equations 11.1 to 11.5. The fundamental dependent variables are here a scalar ρ, a vector u_i and an anti-symmetric tensor F_{ij}.

The left hand side of equation 11.1 may be described as a result of multiplying the scalar ρ by the vector u_i, then differentiating the resulting vector ρu_i, then contracting the tensor so obtained; since the operations of multiplication, differentiation and contraction have been shown to be invariant, the scalar $\partial \rho u_\alpha / \partial x_\alpha$ is independent of the system of coordinates used, and if it is zero in one system of coordinates it is zero in all system of coordinates. The continuity equation expresses, therefore, a fact independent of the system of coordinates employed.

An analogous reasoning applied to 11.3 would show the invariant character of that system. The question of invariance of the first

system of Maxwell's equations requires a special discussion; it can best be treated by introducing a new anti-symmetric tensor D_{ij} connected with F_{ij} by the following relations:

$$F_{12} = D_{34},\ F_{23} = D_{14},\ F_{31} = D_{24},$$
$$F_{34} = D_{12},\ F_{14} = D_{23},\ F_{24} = D_{31}. \tag{11.6}$$

Before we show how this is going to help us in connection with our equations we want to prove that these relations are independent of the coordinate system; i.e., that if 11.6 hold relations of the same form, namely

$$F'_{12} = D'_{34}, \quad F'_{23} = D'_{14}, \quad \text{etc.}, \tag{11.6'}$$

will hold in any other coordinate system. Again, since a general transformation of coordinates can be achieved in steps it will be enough to test a x_1x_2 rotation only. As a result of such a rotation F'_{12} becomes (comp. 9.21)

$$F'_{12} = -F_{11}cs + F_{12}c^2 - F_{21}s^2 + F_{22}sc;$$

using the fact that F_{ij} is anti-symmetric (4.4) we find

$$F'_{12} = F_{12}, \tag{11.7}$$

and since obviously $D'_{34} = D_{34}$ because the x_3x_4 axes are not affected we see that the first of the relations 11.6' follows from 11.6. In order to find F'_{23} we have, according to the general rule following 9.81, to substitute δ_{2i} and δ_{3i} for x'_i and y'_i respectively in $F'_{\rho\sigma}x'_\rho y'_\sigma = F_{\rho\sigma}x_\rho y_\sigma$. As the corresponding values of x_i and y_i we find with the aid of 9.2 considering that $x'_3 = x_3$, $x'_4 = x_4$

$$x_1 = -s, \quad x_2 = c, \quad x_3 = 0, \quad x_4 = 0,$$
$$y_1 = 0, \quad y_2 = 0, \quad y_3 = 1, \quad y_4 = 0;$$

so that

$$F'_{23} = -sF_{13} + cF_{23}; \tag{11.71}$$

in a similar way we obtain

$$D'_{14} = D_{14}c - D_{24}s; \tag{11.72}$$

taking again into account the anti-symmetric property of F we come to the conclusion that the second relation of 11.6' is a consequence of 11.6, and since the same reasoning applies to the remaining relations we conclude that the relations 11.6 are independent of the coordinate system; it is easy to see that they assign to every tensor F_{ij} a tensor D_{ij} (the tensor D_{ij}, or, rather $\sqrt{-1}\,D_{ij}$ is often referred to as the dual of F_{ij}). Now if we express, using 11.6 the F_{ij} in 11.2 in terms of the D_{ij} that set becomes

$$\partial D_{i\alpha}/\partial x_\alpha = 0, \tag{11.2'}$$

and its invariance follows from general considerations as in the case of 11.3.

Formula 11.5 contains only multiplications, contractions and additions, so that there is no doubt concerning its invariance, but the situation changes when we come to the set 11.4. The vector $\partial T_{i\alpha}/\partial x_\alpha$ has been obtained by invariant operations but 11.4 states that only three of its components are zero, a statement which obviously depends on the choice of coordinate axes and is <u>not</u> <u>invariant</u>.

We have now two courses open before us: one is that of resignation, we can say: we see that physics is <u>not</u> like geometry in this respect, that we can only use four-dimensional notations, a four-dimensional diagram but not four-dimensional geometry; the other course is that of adventure, we may try to play the game of geometry; let us pretend that we can apply the formulas of transformation of coordinates in this case; we know that there will be a difference between the theory we obtain and the physics which we undertook to translate into our language; but it may be that the difference will amount numerically to very little. Consider the fourth component of the vector $\partial T_{i\alpha}/\partial x_\alpha$; we found (comp. remark following 6.3) that one of the terms of this expression, $\partial M_{i\alpha}/\partial x_\alpha$ vanishes, and the other $\partial E_{i\alpha}/\partial x_\alpha$ gives $Xu + Yv + Zw$, where u,v,w are the components of velocity, but in order to present the Maxwell equations in a simple form we had to choose our units in such a way that the velocity of light is unity; ordinary velocities are of the order of magnitude of one ten-millionth of the velocity of light, so that we see that by setting the fourth component of $\partial T_{i\alpha}/\partial x_\alpha$ equal to zero we would commit an error that is numerically very small. This encourages us to go on with our adventure and try to force the geometrical character on physics. In order to do that let us go beyond the formal structure of our formulas and recall what the meaning of our fundamental quantities was. The components of the vector u_i were given (see 1.2, 2.4, 4.9) as

$$u_1 = dx/dt, \quad u_2 = dy/dt, \quad u_3 = dz/dt,$$
$$u_4 = i. \tag{11.8}$$

But this identification is obviously not independent of the coordinate system, it gives preference to the fourth coordinate. We may think that this is the source of our difficulty, and that this difficulty may be overcome if we find an invariant identification to take the place of 11.8. The next section will prepare the way for this.

12. Curves in the New Geometry.

The root of the difficulty is that our description of motion was not invariant; motion was described by giving the dependence of the coordinates x,y,z on time, that means by giving three of our coordinates x_1, x_2, x_3, as functions of the fourth x_4 which thus is given preference. The situation is analogous to that in plane analytic geometry where we give y as a function of x, or that in solid analytic geometry when we give y and z as functions of x; in both cases we represent curves; from our four-dimensional point of view we should then consider motion of a particle as a curve in four dimensions (using the word curve in a general sense so that straight line is a special case).

What we want then is a representation of curves in four dimensions which would not give preference to the fourth coordinate. We begin by considering representations of curves in two and three dimensions which give no preference to one coordinate. In the plane a line may be represented by

$$x = ap + b, \qquad y = cp + d,$$

a circle by

$$x = r \cos p, \qquad y = r \sin p;$$

in space a line by

$$x = ap + b, \qquad y = cp + d, \qquad z = ep + f,$$

a helix by

$$x = r \cos p, \qquad y = r \sin p, \qquad z = kp, \text{ etc.}$$

In all these cases to every value of the "parameter" p corresponds a point of the curve; in general, if we set

$$x = f(p), \qquad y = g(p), \qquad z = h(p)$$

we have what we call a parametric representation of a curve (comp. parametric form of equation of straight line in 7.1). In the same way we may represent a curve in four dimensions, which we take to mean motion of a particle, by giving x_i as functions of a parameter p.

The defect of this method is that it contains a certain arbitrariness; we may substitute for p another parameter q by making p an arbitrary increasing function of q. We want now to standardize our parametric representation. The usual way is to choose the arc length along the curve as the parameter. Without going into detail we shall state that arc length between points corresponding to values p_1 and p_2 of the parameter is given by

$$12.1 \qquad s = \int_{p_1}^{p_2} \sqrt{(dx/dp)^2 + (dy/dp)^2 + (dz/dp)^2}\, dp;$$

in the special case when s is used as parameter p we differentiate both sides and obtain

$$12.2 \qquad (dx/dp)^2 + (dy/dp)^2 + (dz/dp)^2 = 1.$$

We may consider in general dx/dp, dy/dp, dz/dp as the components of a vector tangent to the curve; the change of parameter would multiply these derivatives by the same number, i.e., substitute another tangent vector for that one; the quantity $(dx/dp)^2 + (dy/dp)^2 + (dz/dp)^2$ gives the square of the length of the tangent vector; the above equality 12.11 expresses then the fact that if we use arc length as the parameter the length of the tangent vector whose components are the derivatives of the coordinates with respect to the parameter is unity.

We come thus to the idea of a unit tangent vector; it characterizes in every point the direction of the curve; its components are the direction cosines of the tangent.

We may try to go through an analogous process in the case of curves in four-dimensional space, which as we saw may be taken to represent motions; if we succeed, the vector at which we arrive will suggest itself as a natural thing to identify with the vector of components u_i which appears in our formulas. Starting with any parametric representation $x_i = x_i(p)$, where p may be for example t, we try to change our parameter by introducing a new variable q and making p a function of q, choosing this function in such a way that

$$(dx_1/dq)^2 + (dx_2/dq)^2 + (dx_3/dq)^2 + (dx_4/dq)^2 = 1.$$

but $dx_i/dq = (dx_i/dp).(dp/dq)$; so that the function p(q) must be such that

$$dp/dq = \frac{1}{\sqrt{(dx_1/dp)^2+(dx_2/dp)^2+(dx_3/dp)^2+(dx_4/dp)^2}}.$$

Is this possible? In the case when the original p is t the expression under the radical sign will be $u^2 + v^2 + w^2 - 1$, and for motions whose velocities are smaller than the velocity of light (Section 4) this is negative, so that we would get an imaginary value for dp/dq. In order to avoid this unpleasantness we decide to standardize our parameter by requiring $(dx_\alpha/dp).(dx_\alpha/dp)$ to be -1 instead of 1; in this case we find for dq/dt the expression $\sqrt{1 - (u^2 + v^2 + w^2)}$ and we may write

$$12.3 \qquad dx_1/dq = \frac{u}{\sqrt{1-\beta^2}}, \quad dx_2/dq = \frac{v}{\sqrt{1-\beta^2}}, \quad dx_3/dq = \frac{w}{\sqrt{1-\beta^2}}, \quad dx_4/dq = \frac{i}{\sqrt{1-\beta^2}}$$

where β stands for $\sqrt{u^2 + v^2 + w^2}$, i.e., for

what we call speed (the length of the velocity vector). The quantities just written out we want to identify with the components of the vector u_i which appears in our formulas. Since in ordinary cases β is very small, the radical $\sqrt{1-\beta^2}$ is very near to unity and our new identification differs from our old identification (11.8) numerically very little. On the other hand the new values for u_i are according to their derivation the components of a vector, so that if we adopt this identification and also agree to set the fourth component of the divergence of the tensor T_{ij} equal to zero we obtain an invariant theory whose statements differ only very slightly from those accepted in classical physics. It remains to be seen whether there are cases in which the discrepancy is large enough to be tested by experiment.

Chapter III.

SPECIAL RELATIVITY

Guided by the point of view that the formulas of physics ought to be interpreted in four-dimensional geometry we were led to the interpretation of the motion of a particle as a curve in space-time. Following the analogy with a curve in ordinary space where arc length, s, is often used as a parameter, we have introduced a standard parameter, which we may also call arc length and denote by s, for curves in space-time. The partial derivatives dx_i/ds of the coordinates of a point on the curve with respect to s may be considered as the components of a vector tangent to the curve (the square of this vector is -1 in every point - we shall refer to such a vector also as a unit vector). We have then at every point x_1, x_2, x_3, x_4 of such a curve a unit vector dx_i/ds, and we have agreed to identify this vector with the vector u_i which appears in our fundamental laws of physics (11.1 to 11.5) so that

$$u_i = dx_i/ds.$$

In this chapter we want to consider some consequence of this identification.

13. Equations of Motion.

The one thing that was not satisfactory about the formulas of physics was the fact that according to 11.4 only three components of the vector $\partial T_{i\alpha}/\partial x_\alpha$ are equal to zero. In this section we shall see how this defect is corrected by the adoption of the new identification. But before we do that we have to study some immediate consequences of this identification.

Before, a motion of a particle was given by giving the position of the particle in different moments of time, i.e., the coordinates x,y,z as functions of t. Given these functions we can calculate for every moment the velocity vector of the particle - a vector of components

$$u = dx/dt, \qquad v = dy/dt, \qquad w = dz/dt.$$

Now, the same motion is described by giving x_1, x_2, x_3, x_4 as functions of s; and we have a vector of components $u_i = dx_i/ds$ which, due to the special choice of the parameter satisfies the equation

$$13.1 \qquad u_1^2 + u_2^2 + u_3^2 + u_4^2 = -1.$$

Of course, we have merely two representations of the same thing. Given the $x_i(s)$ we can express s as a function of t from

$$x_4(s) = it$$

and substituting the expression of s so found into $x_1(s)$, $x_2(s)$, $x_3(s)$ we will have x,y,z, as functions of t. Or given x,y,z as functions of t we can arrive at the representation $x_i(s)$ as indicated in Section 12.

Also the space-time vector u_i and the space vector u,v,w describe the same thing. The formulas 12.3 show how to find the components u_i in terms of the velocity vector u, v,w, and it is easy to find the u,v,w in terms of components u_i. We simply have

$$u = dx/dt = \frac{dx/ds}{dt/ds} = i\frac{dx_1/ds}{dx_4/ds} = i.u_1/u_4$$

and in a similar fashion

$$13.2 \qquad v = i.u_2/u_4, \qquad w = i.u_3/u_4.$$

We see thus that the vector u_i determines the velocity of motion, and we agree to call it the four-dimensional velocity vector. On the other hand, being a unit vector this vector characterizes the direction in four dimensions of the curve representing motion; its components u_i may be considered as the direction cosines of the tangent (compare Section 7, between formulas 7.41 and 7.5).

But a velocity vector does not characterize the motion of a particle completely; it gives only the kinematical characterization; in dynamics we need in addition, to know the mass of the particle, and then we form the momentum vector (compare beginning of Section 1) whose components are mu, mv, mw. By analogy we form the expressions mu_i or ρu_i (depending on whether we use the discrete or the continuous picture of matter) and consider them as the components of the four-dimensional momentum vector. Using the formulas 12.3 we have for its components

$$13.3 \qquad mu_1 = \frac{mu}{\sqrt{1-\beta^2}}, \qquad mu_2 = \frac{mv}{\sqrt{1-\beta^2}},$$

$$mu_3 = \frac{mw}{\sqrt{1-\beta^2}}, \qquad mu_4 = \frac{mi}{\sqrt{1-\beta^2}}.$$

These are what we call the mathematical components of the momentum vector; its "physical components" are

$$13.31 \qquad a = \frac{mu}{\sqrt{1-\beta^2}}, \qquad b = \frac{mv}{\sqrt{1-\beta^2}},$$

$$c = \frac{mw}{\sqrt{1-\beta^2}}, \qquad d = \frac{m}{\sqrt{1-\beta^2}}.$$

We obtain a relation between the momentum components and mass if we take the sum of the squares of the components 13.3 or 13.31 and use 13.1, viz.,

$$a^2 + b^2 + c^2 - d^2 = -m^2;$$

in words, the negative square of mass is the square of the momentum vector, so that mass is essentially given by the length of the momentum vector; we see here another advantage of the four-dimensional representation: the four dynamical quantities of a particle which in classical physics are given by the three momentum components and mass are here represented more naturally by the four components of a vector.

As stated many times before, numerically $\sqrt{1-\beta^2}$ is in most applications very close to unity so that approximately the first three components of the four-dimensional momentum vector are equal to the components of the three-dimensional momentum vector and the last (physical) component of the four-dimensional momentum vector is, in first approximation, equal to mass.

Let us consider more in detail this fourth component of the momentum vector. If we want a better approximation we develop the last of 13.31 according to powers of β and keep only two terms; we have thus the approximate equality

$$13.4 \qquad d = \frac{m}{\sqrt{1-\beta^2}} \doteq m + \tfrac{1}{2}m\beta^2;$$

the correction represented by the second term is nothing but kinetic energy; of course, if ordinary units are used this term has to be written

$$13.41 \qquad \tfrac{1}{2}mV^2/c^2$$

where c is the velocity of light, and V is the velocity of the particle measured in the same units, because β is the ratio of the velocity of the particle to that of light. We had better say then that the correction is kinetic energy divided by the square of the velocity of light. Sometimes this fact is expressed by saying (neglecting the other terms, which are very, very small) that when a body is in motion its mass is increased by its kinetic energy (divided by the square of the velocity of light).

The interest of this lies in the close relationship which is thus established between mass and energy - a relationship that plays a prominent part in present physics.

Sometimes the whole expression $m/\sqrt{1-\beta^2}$ is referred to as energy of the particle; mass, from this point of view, appears then as part of the energy, that part that the particle possesses even when it is at rest; in other words, mass appears as the rest-energy of the particle. We could also call $m/\sqrt{1-\beta^2}$ generalized mass and say that mass changes as a result of motion (compare end of this section).

We are ready now to discuss the equations of motion 11.4 or

$$13.5 \qquad \partial M_{i\alpha}/\partial x_\alpha = \partial E_{i\alpha}/\partial x_\alpha.$$

The left hand sides may be written as

$$\partial(\rho u_i u_\alpha)/\partial x_\alpha = \partial\rho u_\alpha/\partial x_\alpha . u_i + \rho u_\alpha . \partial u_i/\partial x_\alpha ;$$

the first factor of the first term vanishes according to the continuity equation 11.1; the second term may be written, recalling the definition of u_i, as

$$\rho . \partial u_i/\partial x_\alpha . u_\alpha = \rho . \partial u_i/\partial x_\alpha . dx_\alpha/ds = \rho . du_i/ds;$$

the right hand side of 13.5, according to our former calculation (Section 6) is $\varepsilon F_{i\alpha}u_\alpha$, so that the equations of motion become

$$\rho . du_i/ds = \varepsilon F_{i\alpha}u_\alpha;$$

or, if we use the discrete picture, considering both mass and electric density to be concentrated in one point, and denoting mass by m, electric charge by e,

$$13.51 \qquad m . du_i/ds = eF_{i\alpha} . dx_\alpha/ds.$$

These are the equations we are going to discuss. In applying them to physics we give preference to time by writing

$$13.52 \qquad m . du_i/dt = eF_{i\alpha} . dx_\alpha/dt$$

which spoils the invariant form but does not change the contents of the statement because the transition from 13.51 to 13.52 is equivalent to multiplication by ds/dt. Using 4.72 (or 6.1) the last equations become

$$13.53 \qquad \begin{aligned} m . du_1/dt &= e(X + Nv - Mw), \\ m . du_2/dt &= e(Y + Lw - Nu), \\ m . du_3/dt &= e(Z + Mu - Lv), \\ m . du_4/dt &= ie(Xu + Yv + Zw). \end{aligned}$$

Multiplying the left hand side of the first of these equations by $i.u_1/u_4$ and the right hand side by u (compare 13.2); using in the same way $i.u_2/u_4 = v$ on the second, and $i.u_3/u_4 = w$ on the third, and adding the results we get the fourth equation because the left-hand side comes out

$$(im/u_4).(u_1 . du_1/dt + u_2 . du_2/dt + u_3 . du_3/dt)$$

and differentiating the identity 13.1, we find that the second factor is, $u_4 . du_4/dt$. The fourth equation is thus a consequence of the first three, a great improvement over the situation as it was before the new identification. The fourth equation also has a definite physical

meaning now; the left hand side may be said to represent the time rate of change of energy (since the variable part of mu_4 has been recognized as kinetic energy), and the right hand side has been recognized before (Section 6) as the rate at which the energy of the field (potential energy) is being expended in moving the body. The difficulty with the fourth equation has thus been settled in a most satisfactory fashion but the system as a whole, or the first three equations, have to be tested by experiment (the fourth, being a consequence of the first three, cannot be wrong if these three will be proved to be "true").

Since

$$du_i/dt = \frac{d}{dt}(dx_i/ds) = dt/ds.\frac{d}{dt}(dx_i/dt) = \frac{1}{\sqrt{1-\beta^2}} \cdot d\dot{x}_i/dt^2$$

we may write our equations as

$$m'.d^2x/dt^2=e(X+Nv-Mw), \quad m'.d^2y/dt^2=e(Y+Lw-Nu),$$

13.6

$$m'.d^2z/dt^2 = e(Z + Mu - Lv),$$

where

13.7 $$m' = m/\sqrt{1-\beta^2}.$$

The right hand sides of these equations (as stated in Section 6) are the components of the force exerted by the electromagnetic field on the particle. Comparing the left hand sides with the classical expressions we see then that the correction resulting from our identification is equivalent to the substitution of m' for m in the classical equations of motion. We may say then that our theory predicts that motion will be governed by the old equations in which mass has been replaced by a corrected mass the correction being the kinetic energy (divided by the square of the velocity of light). In the vast majority of cases the factor $1/\sqrt{1-\beta^2}$ is very close to one, but there are a few cases where it is not, and these cases afford an opportunity to test the new theory and to see whether it or the old one is better adapted to give account of experimental results. In experiments with "cathode ray particles" by Bucherer the predictions of the new theory seem to have been verified.

14. Lorentz Transformations.

Now that we saw that the new identification removes the difficulty in connection with the fourth equation of motion we want to consider some other consequences, and in the first place we want to give a discussion of the physical significance of the transformation of coordinates promised in Section 10. The new feature about our coordinate systems is the greater arbitrariness in their choice. Before we were free to pass from one system to another with the same time axis; now we may change the time axis also (formulas 10.6), and we want to see what it means.

In general, in one system of coordinates a geometrical configuration is described by certain numbers (e.g., coordinates of its points) and certain equations (e.g., equations of its straight lines); in another system of coordinates the same configuration will be characterized by other numbers and other equations, but it will be another description of the same configuration; or, we may say, another identification which is theoretically just as good as the first, but may be more, or less, convenient for practical purposes. In general we make our choice of a coordinate system guided by the properties of the object we are studying and our own position in space. If we study an ellipsoid we would choose for coordinate axes its principal axes; or in another case we would choose the direction away from us as the y-axis, the direction to the right as the x-axis, the vertical direction as the z-axis; but in principle all axes are permitted. The same general situation obtains in physics, considered as four-dimensional geometry. We have many systems of coordinate axes at our disposal, and we want to investigate now what use we can make of this arbitrariness, how we can adjust the choice of axes to the requirements of a particular situation. In particular, we are interested in the choice of the t-axis.

The object we want to study in the first place is the motion of a particle. We represent such a motion by a curve in four-space (and a straight line we consider as a special case of a curve). At every point of that curve we have a unit tangent vector, the four-dimensional velocity vector of components a,b,c,d; or we may characterize it by the three-dimensional velocity components u,v,w; if we pass to another coordinate system the components a,b,c,d will be changed, and so will the components u,v,w. If the coordinate transformation affects only the space coordinates x,y,z then the component d will not be affected, and therefore ρ will not change; in other words, u,v,w will be changed but not $u^2 + v^2 + w^2$; the velocity vector will have different components, but its absolute value, the speed, will be the same. This is essentially as in old physics; the new feature is in the existence of transformations affecting t, and the most striking result of it is expressed in the following theorem.

Theorem. For every motion it is possible for every moment to choose a system of space-time coordinates in such a way that the speed be zero.

Proof. Begin with any axes; then, without changing the time axis, change the space axes so that the motion, at the moment considered,

takes place along the z-axis; we have then $a = b = 0$; now consider a transformation involving z and t; denoting the new components of the four-dimensional velocity vector a', b', c', d' we shall have (10.6)

$$a' = 0, \quad b' = 0, \quad c' = c\sigma + d\tau, \quad d' = c\tau + d\sigma;$$

if we want to make $c' = 0$ we have to choose the angle ψ so that

$$\frac{\tau}{\sigma} = \sin \psi = -c/d;$$

if c is in absolute value less than d, and this is so for all motions of bodies so far observed, an angle satisfying this relation and therefore a system of coordinates for which $a = b = c = 0$ can be found. From formulas 13.2 it follows that in such a system $u = v = w = 0$, and the theorem is proved.

The theorem just proved is expressed often by saying that <u>every particle can be transformed to rest</u>.

After we have found a coordinate system in which a particle is at rest we can perform any transformations of space coordinates and the property will not be destroyed; any transformation involving time, on the contrary will result in introducing significant space components of the velocity vector; we see thus that whether a particle is at rest in a coordinate system depends exclusively on the choice of the time axis, so that the choice of the time axis is equivalent to the choice of a body which we desire to consider at rest; in other words, the direction of the time axis may be characterized by indicating what particle is at rest in the corresponding coordinate system.

What time axis we actually choose depends, as in geometry, on circumstances; in many cases we shall want to consider <u>ourselves</u> as being at rest, or our laboratory, or the earth.

In what precedes we spoke of a motion of a particle <u>at a given moment</u>; in a given system of space-time coordinates a particle may be at rest at one moment and not at rest some other time; but there exists a class of particles which if transformed to rest for one moment will be at rest always; these are those particles whose representative four-dimensional curves have the same tangent vector at all points, i.e., are straight lines; it is clear that if the direction of such a straight line is taken as the direction of the t-axis the velocity in the so obtained coordinate system is zero. But if we choose any other (cartesian) coordinate system, the three-dimensional velocity u,v,w, will be constant in absolute value and direction, so that we have a rectilinear uniform motion. From our point of view then the distinction between uniform rectilinear motion and rest is a non-essential distinction, this distinction does not exist until we introduce a coordinate system; it is of the same nature as the distinction between lines which are and those which are not parallel to the x-axis in ordinary analytic geometry.

If a motion is not uniform and rectilinear then there is no coordinate system in which the particle is permanently at rest. But rather than to make a strict distinction between particles which are and those which are not in uniform rectilinear motion or rest, it is more in keeping with our point of view to speak of particles which may be (within experimental error) considered at rest for a sufficiently long period of time.

We are now in a position to explain the name and the origin of the theory we are studying. We saw that in this theory there is no such thing as absolute rest or absolute motion of a body. If it is at rest with respect to one system of coordinates it may move with respect to another and vice versa; we can only speak of relative motion; that is where the name Relativity comes from.

If we adopt this point of view, we have to consider as permissible all transformations of coordinates from one to any other cartesian coordinate system. Later on we shall consider other, more general, systems of coordinates, and therefore more general coordinate transformations; we shall replace our equations by more general ones which will be invariant under these more general coordinate transformations; in comparison with this situation we may say that we consider now only special coordinate transformations and invariance under them; therefore the present theory is called "Special Relativity Theory".

It may be mentioned that the historical order of appearance of the ideas of our subject - as it happens so often - has been quite different from the order which seems natural and in which we have presented them. First the formulas of transformation involving space coordinates and time have been introduced by Lorentz without, however, giving to them the meaning they have now; in Lorentz's theory there exists one universal time t, and other times t' play only an auxiliary part. The merit of making the decisive step and recognizing the fact that all these variables are on the same footing - belongs to Einstein (1905). The four-dimensional point of view, after some preliminary work had been done by Poincare and Marcolongo, was most emphatically introduced by Minkowski in 1908.

15. Addition of Velocities.

As explained in Section 13 we have two ways of characterizing the velocity of a body: by means of the three-dimensional velocity vector and by means of the four-dimensional velocity vector. We can pass from one representation of

velocity to the other without difficulty and the two methods are equivalent as long as we do not change our coordinate system. But if we come to study the relative motion of one body with respect to another and want to define the relative velocity, the four-dimensional point of view leads to conceptions which are at variance with commonly accepted ideas and we want to devote this section to the clarification of this situation. It is natural to reduce the definition of relative velocity of a body with respect to another body to the conception of the velocity of a body in a coordinate system by saying: By velocity of the body B with respect to a body A we mean the velocity of B in a system of coordinates in which the velocity of A is zero.

If we want to find the velocity of B with respect to A we have to transform our coordinates so that in the transformed coordinate system A be at rest. It is clear that the meaning of relative velocity is made to depend by the preceding definition on what we mean by transformation of coordinates. If by transformation of coordinates we mean only transformation of three-dimensional coordinates - transition to moving axes - we have the old idea of relative velocity; if, on the other hand, we consider four-dimensional coordinate axes and our transformation of coordinates involves the coordinate x_4, or t, in the sense of the theorem of Section 14, it is clear that we give a new meaning to relative velocity, and we should not be surprised if the so defined "relativistic" relative velocity will possess properties different from those of the "classical" relative velocity.

Consider a body A and a body B that moves with respect to A uniformly and rectilinearly with a velocity V_{BA}; this means according to our definition that the velocity of B in a coordinate system in which A is at rest is V_{BA}. Introduce a coordinate system in which A is at rest and B moves along the z-axis, and call the coordinates x_A, y_A, z_A, t_A; introduce also a system of coordinates in which B is at rest so that (10.6)

$$15.1 \qquad x_A = x_B, \qquad y_A = y_B,$$
$$z_A = z_B\sigma_{BA} + t_B\tau_{BA}, \quad t_A = z_B\tau_{BA} + t_B\sigma_{BA}.$$

Now describe the motion of the body B in each system neglecting the x and y coordinates. In the system A the motion of the body B which, we assume, at the moment $t_A = 0$ was at $z_A = 0$ is given by

$$z_A = t_A \cdot V_{BA};$$

in the system B the body B is at all times at the origin of the coordinate system, so that $z_B = 0$; substituting this value in the transformation formulas and eliminating t_B we get $z_A = t_A \cdot \frac{\tau_{BA}}{\sigma_{BA}}$. Comparing this with the preceding equation we have (10.5)

$$15.2 \qquad V_{BA} = \frac{\tau_{BA}}{\sigma_{BA}} = \sin \Psi_{BA}.$$

Solving the above transformation formulas for z_B, t_B we also find that $\tau_{AB} = -\tau_{BA}$, $\sigma_{AB} = \sigma_{BA}$ so that

$$V_{AB} = -V_{BA}.$$

This result, that the relative velocity of A with respect to B is the negative of the relative velocity of B with respect to A is in keeping with the old ideas.

Now consider three bodies, A,B,C, all moving in one direction (more precisely B and C moving in the same direction with respect to A). Denote the velocities of B and C respectively with respect to A by V_{BA} and V_{CA}, and the velocity of C with respect to B by V_{CB}. We have in addition to the above transformation formulas the formulas

$$15.11 \qquad z_A = z_C\sigma_{CA} + t_C\tau_{CA}, \quad t_A = z_C\tau_{CA} + t_C\sigma_{CA},$$

and

$$15.12 \qquad z_B = z_C\sigma_{CB} + t_C\tau_{CB}, \quad t_B = z_C\tau_{CB} + t_C\sigma_{CB},$$

and also

$$15.21 \qquad V_{BA} = \frac{\tau_{BA}}{\sigma_{BA}}, \quad V_{CA} = \frac{\tau_{CA}}{\sigma_{CA}}, \quad V_{CB} = \frac{\tau_{CB}}{\sigma_{CB}}.$$

Express now z_A, t_A in terms of z_C, t_C by substituting the values for z_B, t_B given by the transformation formulas 15.12 into the transformation formulas 15.1; comparing the result with the transformation formulas 15.11 we get

$$\sigma_{CA} = \sigma_{CB}\cdot\sigma_{BA} + \tau_{CB}\cdot\tau_{BA}, \qquad \tau_{CA} = \tau_{CB}\cdot\sigma_{BA} + \sigma_{CB}\cdot\tau_{BA}$$

whence, using the above expressions of velocities in terms of transformation coefficients, 15.2 and 15.21,

$$15.3 \qquad V_{CA} = \frac{V_{BA} + V_{CB}}{1 + V_{BA}\cdot V_{CB}}.$$

This is the Einstein formula for addition of velocities for the case of two motions in the same direction. This formula should be compared to the formula of addition of classically defined relative velocities

$$15.4 \qquad V = V' + V''.$$

Of course, there is no contradiction between the two formulas because they refer to different quantities. Still it is legitimate to ask which formula is better from the point of view of experiment, which - if any - is "correct"

for the relative velocities that we actually measure.

In ordinary units the second term in the denominator in formula 15.3 should be divided by the square of the velocity of light, so that for moderate velocities the formulas give results that differ numerically very little, and it seems to be difficult to devise an experiment with high enough velocities of material particles so that the formulas could be tested directly. In the next section we shall consider the case when one of the velocities is that of light; in the meantime we may mention that formula 15.3 is a special case of a more general formula which corresponds to the case when the two motions are not in the same directions. This general formula gained temporary importance some years ago when it played a decisive role in the early stages of the application of the idea of the spinning electron to the explanation of spectra.

16. Light Corpuscles, or Photons.

In studying curves in four-space representing motions of particles we succeeded (Section 12) in choosing a standard parameter, s, by considering the expression

$$(dx_1/dp)^2 + (dx_2/dp)^2 + (dx_3/dp)^2 + (dx_4/dp)^2$$

and by setting ds/dp equal to the reciprocal of the square root of minus the above expression. This procedure would not work if the above expression were equal to zero. We can imagine in our four-dimensional geometry curves and straight lines for which the above expression *is* zero (Section 10), and the question arises: what will be the physical interpretation of such curves; in other words: is there anything in physics that could be identified with such curves in the same way that motions of particles are identified with curves for which the above expression is negative. In order to answer this question let us calculate the three-dimensional velocity corresponding to such a curve; if the above expression is zero for one choice of parameter it will be zero for all choices; using t as parameter, and using physical coordinates we have then

$$(dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2 - 1 = 0$$

$$\text{or} \quad u^2 + v^2 + w^2 = 1,$$

i.e., we can say that the curves of zero square tangent vectors correspond to what we have to call from the three-dimensional point of view *particles moving with the velocity of light*. This suggests to identify such curves in some way with propagation of light.

Since the time of Newton and Huygens two theories of light have been vying for supremacy with variable success; according to one, the so-called corpuscular theory, light consists (like matter on the discrete theory) of particles which lately (Wolfers, 1923) have been called "photons"; according to the other theory light is a wave phenomenon. For our present purposes the former view seems to be better adapted. If we adopt it we can make our former statement more specific by saying that *we identify curves of zero square tangent vectors with photons, or with motion of photons.*

In adopting thus the corpuscular theory of light we do not in the least mean to say that the corpuscular theory of light is correct, and still less that the other theory - the wave theory - is wrong. We simply want to show that the identification just mentioned permits us to give account of certain light phenomena; and it is enough to mention polarization in order to see that other phenomena are left out.

To begin with we want to point out an advantage that the Relativity theory has compared to classical theory in the matter of corpuscular theory of light. In classical theory difference in velocity is merely a *quantitative* difference, in relativity this means an entirely different *kind* of curves, and there are other differences entirely of *qualitative* nature that are consequences of our identification, which is more in keeping with the nature of light compared to matter as we know it from experiment. This seems to constitute a very strong argument in favor of the adoption of the point of view of Relativity in general, and of the identification we are discussing now in particular.

We want next to discuss what is usually referred to as constancy of the velocity of light. The reader may have noticed that a while ago when we were calculating the three-dimensional velocity, corresponding to curves with zero-square tangent vectors, we did not say in what coordinate system we wanted to calculate this three-dimensional velocity. As a matter of fact, the result shows that it is independent of the coordinate system; i.e., no matter what bodies we consider as being at rest, we come out with the same value for the velocity of light, in our units - one.

This seems surprising; it contradicts the commonly accepted ideas concerning addition of velocities; but we have been led to a different formula for the addition of velocities, and we can show that the constancy of velocity of light is in agreement with that formula 15.3. In fact, if we consider the case that C moves with the velocity of light (that is *one* in our units) with respect to B, that means that $V_{CB} = 1$; substituting this value in 15.3 we find that V_{CA} is also one; that is, what is motion with velocity one in one system is motion with velocity one in another system.

This discussion, of course, proves nothing but the inner consistency of the theory.

Another question is whether constancy of velocity of light, i.e., independence of this velocity from the choice of the system which is considered at rest is consistent with experiment. As a matter of fact, it appears that it is; the weight of experimental evidence seems to be for it. Historically, results of some experiments by Michelson and Moreley performed in 1887 and pointing in the same direction played a great role in the creation of the Theory of Relativity.

Having considered thus the question of velocity of light we pass to the discussion of another consequence of our identification.

We have decided in a general way to identify straight lines whose vectors are of zero square with light or the motion of photons in the same way that straight lines whose vectors are of negative square are identified with matter or uniform rectilinear motion of material particles. But a straight line (in four dimensions) does not characterize the motion of the particle completely - it only gives the velocity of the motion, it characterizes it only kinematically; for a complete dynamical characterization we had to introduce (Section 13) the mass of the particle, and that led us to introduce the momentum vector, whose square we found to be $-m^2$; the complete characterization of a material particle consists then of a line with a vector (of negative square) on that line. In the same way we shall characterize the motion of a photon by a line with a vector of zero square on it. We have thus the same picture for a material particle and a photon; in both cases we have a line with a vector on it; only in the first case it is a vector of negative square; in the second of zero square; this difference corresponds to the difference in the speeds of the particles in the classical theory. But in the classical theory this is a purely quantitative difference and here, as mentioned before, it leads to qualitative differences, some of which we are going to consider.

In the first place a photon cannot be transformed to rest. In fact transforming a photon to rest would mean finding a coordinate system such that in it the time axis will have the direction of the photon; but that would mean that the vector 0,0,0,1 would have zero square which is impossible.

Then there is this distinction: two material particles may differ in mass, that means in the squares of their momentum vectors, and this is an essential difference because the square of a vector is not affected by a transformation of coordinates; all photons on the other hand have vectors of the same square, namely zero. We shall prove that as a consequence of this, two photons never differ essentially, that is, given two photons there always exist two systems of coordinates in which the descriptions of the two photons are the same. To begin with, we may choose the origins of the two coordinate systems on the respective straight lines; next we may consider the two lines in the respective z-t planes. The momentum vectors of the corresponding photons will have now in their respective coordinate systems the components $0,0,q_3,q_4$ and $0,0,p_3,p_4$ (contrary to our general agreement we use here subscripts with physical coordinates) and since both vectors are of zero square we'll have

$$q_3^2 - q_4^2 = 0, \qquad p_3^2 - p_4^2 = 0;$$

by choosing appropriately the sense on each coordinate axis we can reduce these conditions to

16.1 $$q_3 = q_4, \qquad p_3 = p_4.$$

Now perform in the second system the transformation 10.6

$$z' = z\sigma + t\tau, \qquad t' = z\tau + t\sigma,$$

which applied to the second vector and taking into account 16.1 gives

16.2 $$p_3' = p_4' = p_3(\sigma + \tau).$$

But σ and τ are subject only to the condition that $\sigma^2 - \tau^2 = 1$, so that we can choose $\sigma + \tau$ arbitrarily; if we make the choice

16.3 $$\sigma + \tau = q_3/p_3,$$

we shall have $p_3' = q_3$ and the statement is proved.

This theoretical conclusion, that any two photons are not essentially different from each other must be confronted with experience. At first sight it seems to contradict it. We know that light differs from case to case; it differs in intensity and color. For difference in intensity we account by assuming that every beam of light consists of many photons so that intensity (for a given color) is proportional to the number of photons in the beam. Remains color. But experiments show that color actually depends on the state of motion of the observer; when an observer approaches a source of light, color seemingly changes (Doppler effect) and so the field is clear for our assertion. Now let us see how it works out.

Before we treat the situation from the point of view of the Relativity theory we have to say a few words about how color appears in physics as a measurable quantity. From the point of view of the wave theory to light is attached a certain measurable quantity ν - "frequency" which corresponds to color in the sense that different colors correspond to different frequencies. On the corpuscular theory photons are characterized by their energies, E, and the fundamental relation between frequency

and energy is given by the formula

16.4 $$E = h\nu,$$

where h is the so-called Planck's quantum constant, which for us appears simply as coefficient of proportionality establishing the relation between the values of two quantities which measure the same thing in different units, much in the same way that c, the velocity of light, appears in the formula connecting mass and energy (compare 13.41). Of the two quantities, E and ν, which can be used to measure color, E will be the one that is more convenient for our purposes because we use the corpuscular theory of light.

The question now is with what quantity in our theory are we going to identify E. In order to have a suggestion we notice that E is of the character of kinetic energy; it plays for light particles the same role that kinetic energy plays for material particles. There (Section 13) we identified kinetic energy with the second term in the development

$$\frac{m}{\sqrt{1-\beta^2}} = m + \tfrac{1}{2}m\beta^2 + \ldots$$

of the fourth component of the momentum vector; of the other terms the third and the following are negligible for material particles, and the first is a constant so that it plays no part in these considerations where only differences in energy are important; besides, the corresponding constant for light is zero; everything leads us thus to compare E with the fourth component of the momentum vector of light, or photon. We arrive in this way to a <u>new identification</u>: we identify the mathematical quantity "time component of the momentum vector of a photon" with the physical quantity E which, except for a factor of proportionality, is frequency and measures color. This identification makes color dependent on the coordinate system but this dependence, as was said before, is to be expected, and our next question is whether the character of this dependence corresponds to experimental facts.

Suppose that E is the energy of a light corpuscle in one system of coordinates; what will it be in another? We have already calculated how the components of a zero square vector change under a transformation of coordinates involving time. Formula 16.3 shows that the ratio of the fourth components of the two vectors, and according to our identification this means the ratio of the energies or frequencies, is

$$\nu'/\nu = \sigma + \tau .$$

On the other hand, we saw before (15.2) that the relative velocity of two bodies which are at rest in the corresponding systems of coordinates is

$$V = \tau/\sigma ;$$

this gives

$$\frac{1+V}{1-V} = \frac{\sigma+\tau}{\sigma-\tau}$$

and taking into account the identity $\sigma^2 - \tau^2 = 1$ we find

16.4 $$\nu'/\nu = \sigma + \tau = \sqrt{\frac{1+V}{1-V}},$$

or, in first approximation $1 + V$.

Let us try to figure out the predicted change of frequency on the classical (wave) theory. If we have a wave of frequency ν that means that there are ν vibrations per unit of time, and since the velocity is unity, there will be ν waves per unit of length. Now, if we move toward the source with a velocity V we shall travel in a unit of time the distance V and we shall meet $V.\nu$ additional vibrations, so that the number of vibrations our eye receives in a unit of time will be $(1 + V).\nu$, and this will be the frequency for the moving observer. The two theories give then the predictions

$$\sqrt{\frac{1+V}{1-V}}, \qquad 1 + V$$

for the change in frequency due to motion of the observer, and the difference between these two values is too small to be subjected to an experimental test; within experimental error both seem to fit observations equally well.

17. Electricity and Magnetism in Special Relativity.

In the preceding sections of this chapter we have discussed some modifications that are brought about by the Theory of Relativity in Kinematics, Mechanics, and Optics. There are other modifications which have attracted a great deal of popular attention due to their sensational and paradoxical character. We shall only mention the so-called effects of motion on the shape of bodies, lengths, measure of time, and the fact that in the Theory of Relativity the conception of simultaneity loses its absolute character so that two events which are to be considered simultaneous in one system of space-time coordinates need not be simultaneous in another. But we shall say a few words about electricity and magnetism. Even in the first chapter the components of the electric and the magnetic force vectors were combined into one tensor F_{ij}, so that electricity and magnetism seem to be treated as two aspects, or manifestations of a higher entity. But as long as we limit

ourselves to transformations of space coordinates the components of F corresponding to electricity are transformed among themselves and those corresponding to magnetism - among themselves, so that their unification in one tensor F_{ij} may be considered as artificial. However, when we introduce transformations of space-time coordinates (formulas 10.6) the situation changes radically.

Following the procedure used in Section 11 when we were proving the invariant character of the relations between the tensors F and D we can deduce the following formulas corresponding to rotation in the x_3x_4 plane.

$$F'_{41} = F_{31}s + F_{41}c \qquad F'_{23} = F_{23}c - F_{24}s$$

$$F'_{42} = F_{32}s + F_{42}c \qquad F'_{31} = F_{31}c - F_{41}s$$

$$F'_{43} = F_{43} \qquad F'_{12} = F_{12}.$$

From these mathematical formulas we can pass to formulas involving physical components and only real quantities by making use of 4.72, 10.3 and the fact that $F_{ij} = -F_{ji}$. We obtain thus the relations

$$X' = \sigma X + \tau M \qquad L' = \sigma L - \tau Y.$$

$$Y' = \sigma Y - \tau L \qquad M' = \sigma M + \tau X$$

$$Z' = Z \qquad L' = L.$$

$$N' = N$$

The interpretation of these formulas is that if the unprimed letters give the components of electric and magnetic force in one system the primed letters will give the components of electric and magnetic force in a system which moves with respect to the first with velocity $V = \tau/\sigma$.

These formulas show that the distinction between electric and magnetic forces is not an absolute distinction, but depends on the coordinate system used; we might, for example, have in the old coordinate system a purely electric field, $L = M = N = 0$; in the new system the magnetic components will be different from zero, viz., $-Y_\tau$, X_τ, 0. What is the physical meaning of this? It means that a field may have electric effects on one body but electric and magnetic effects on a body that moves with respect to it. This prediction is verified by experiment. The fact may be restated by saying that an electric charge in motion has magnetic effects, it may, e.g., deflect a magnetic needle. As an example, the magnetic field of a moving electron may be easily calculated. We start with an electron at rest. Its magnetic field is supposed to be zero, its electric field is supposed to be independent of time and symmetric with respect to the point; under these conditions Maxwell's equations reduce, as can be seen easily, to Newton's equations which, as we know, give the inverse square law for the electric forces. The field of an electron in motion can now be obtained by applying the above formulas.

Chapter IV.

CURVED SPACE

The theory that has been developed so far may be said to consist of two parts: a general part which may be called Geometry and which, in addition to material analogous to that treated in ordinary geometry, includes general rules of operations on tensors, and a special part which may be called Physics and which deals with three definite tensor fields, a scalar field ρ, a vector field u_i, an antisymmetric tensor field F_{ij}, which all have been combined into the tensor field T_{ij}, and with special conditions which we impose on these fields, viz., equations 11.1 to 11.5. The second part is independent of the first in the sense that we could have built with the same geometry a different "physics", we could have chosen another set of tensors instead of ρ, u_i, F_{ij}. The reason why our physics was independent of our geometry is because the latter does not furnish us any tensors, except the tensor δ_{ij}, or the tensor of scalar multiplication which is, so to say, the same in all points (and at all times) and therefore cannot be used to explain the variety and change which are characteristic of the outside world. In other words, our geometry does not possess any structure which seems to be necessary for the interpretation of the outside world and therefore we had to superimpose on our geometry a certain arbitrary structure by introducing special tensors, by filling, so to say, the empty space-time with these tensor fields. Our geometry does not give us a landscape, it gives us, so to say, only a frame for one, or only a stage and the landscape can be constructed on it by means of stage-settings which do not constitute an organic part of the stage. Although some success has been achieved with the theory just described we may want to accomplish more, we may want to have a geometry possessing a structure of its own which might be used in interpreting the outside world. Such a possibility is suggested by the consideration of curved surfaces. The space-time we have been working with is of the same character as a plane (except for the number of dimensions) it is as devoid of structure as a plane. A curved surface, on the other hand, possesses a certain structure; it is not necessarily the same in all points, there may be a difference in curvature. We shall investigate the possibility of a four-dimensional space which bears the same relationship to our flat space-time as a curved surface bears to a plane; we shall expect to find that it possesses a certain structure which we'll try to interpret in terms of our physical quantities; more specifically, since all our physical quantities have been combined (by formula 11.5) into a symmetric tensor of the second rank, viz., T_{ij}, we shall expect to find a tensor of that character connected with the curvature of our curved four-dimensional space.

The plan of our study will be to begin with the simplest case, a case that is even simpler than that of a surface, viz., with the case of a curve, and then to work up gradually.

18. Curvature of Curves and Surfaces.

We consider a curve in the plane. We assume that it possesses a tangent at every point, and, furthermore, that if the origin of coordinates is chosen in any point of the curve, and the tangent at the origin is chosen as the x-axis, the curve, in the neighborhood of the origin may be represented by a function

$$y = f(x),$$

which can be expanded into a power series converging in the neighborhood of the origin. The constant term of this expansion vanishes because the curve passes through the origin so that the equation must be satisfied for $x = 0$, $y = 0$; the coefficient of the first power of x also vanishes, since the slope of the tangent, which is the x-axis, must be zero; if we write the next term in the form

$$\tfrac{1}{2}a_2x^2$$

the coefficient a_2 is called the curvature of curve at the point considered, i.e., the point chosen for the origin. Since every point can be chosen for the origin this assigns to every point of the curve a curvature. We may say that if we drop all terms of the expansion following the one just written out, i.e., consider the curve

18.1 $$Y = \tfrac{1}{2}a_2x^2$$

this curve (a parabola) is an approximation to the given curve in the neighborhood of the point considered.

We consider next a surface. Here we assume that it has a tangent plane at every point. Taking a particular point on the surface as the origin and the tangent plane at that point as the xy-plane we may represent the surface by an equation $z = f(x,y)$. We again assume that for every point on the surface this function may be developed into a power series converging in the neighborhood of that point. The constant term and the coefficients of the first powers of the variables will vanish as before. We write out the next group of terms, those that are quadratic in the variables, in the form

18.2 $Z = \frac{1}{2}(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2)$,

where we use x_1 for x, and x_2 for y.

We may consider the coordinates x_1, x_2 as the components of a vector in the tangent plane which joins the origin to the projection of the point on the surface, whose coordinates are x_1, x_2, $z = f(x_1,x_2)$. The expression 18.2 assigns thus to every vector in the tangent plane a number (which may be considered as the ordinate of a paraboloid approximating the surface). This assignment is independent of the coordinate system, i.e., if we choose another system of coordinates we shall have the same number assigned to the same vector although its components will have changed; in fact in a rotation of the coordinate axes the degree of a polynomial is not affected so that the group of second degree terms in the expansion of z is transformed into the group of second degree terms of that expansion in the new coordinates. We have then a function whose values are numbers and whose argument is a vector; is it a tensor? Of course not; but it is easy to introduce a tensor with which our function 18.2 is closely connected. We simply write, as in 9.1

18.3 $a_{11}x_1y_1 + a_{12}x_1y_2 + a_{21}x_2y_1 + a_{22}x_2y_2 = s(x,y)$

using the coefficients a_{11}, a_{12}, a_{22} of our function 18.2 and writing in the third term a_{21} for a_{12} for the sake of symmetry; this is a (symmetric) tensor of the second rank depending on two vector arguments x and y, and from which our function is obtained by setting the vector arguments equal to each other. We arrived thus in the case of a surface at a tensor of rank two which expresses the curvature properties, i.e., the structure of the surface insofar as it describes its deviation from its tangent plane in the neighborhood of the point of contact. This encourages us in our enterprise: if we succeed in generalizing this result to higher dimensions, we may try to identify the generalization of this symmetric tensor of rank two with the symmetric tensor of rank two which, as we saw, combines in itself matter and electricity. We want to state at this time that we shall be ultimately successful in our enterprise but that everything will not run very smoothly, and we shall have to make an effort in order to arrive in the general case at a tensor of rank two. The configurations which will present themselves immediately will not be exactly tensors, and even after we shall arrive at a tensor it will not be a tensor of rank two. We shall have to overcome these obstacles, and in order to be able to do that we shall need some preparation, which we shall make by studying more attentively the case of a surface before we pass to the consideration of more complicated cases.

The curvature of a curve is characterized by a number; that of a surface is a more complicated thing and is characterized by a tensor; we know, however, that there are certain numbers connected with a tensor, in an intrinsic way (that is, independent of the system of coordinates), viz., the numbers given by 9.5 and 9.51, and we may expect that they have geometrical significance. In fact the first one, $a_{11} + a_{22}$, is known as the mean curvature, and the second,

18.4 $$K = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}$$

as the <u>total curvature</u> of the surface at the point considered. We know that K is independent of the choice of a system of coordinates, but we want to show how it can be obtained without the use of any coordinates at all. We have introduced above a vector notation $s(x,y)$ for our tensor 18.3; we now write out the expression (the expressions in coordinates are written down for the sake of future references and may be disregarded at present):

18.5 $$\begin{vmatrix} s(x,u) & s(x,v) \\ s(y,u) & s(y,v) \end{vmatrix} = \begin{vmatrix} a_{\alpha\gamma}x_\alpha u_\gamma & a_{\alpha\delta}x_\alpha v_\delta \\ a_{\beta\gamma}y_\beta u_\gamma & a_{\beta\delta}y_\beta v_\delta \end{vmatrix},$$

where x,y,u,v are arbitrary vectors, and we assert that it is equal to

18.6 $$K.\begin{vmatrix} x.u & x.v \\ y.u & y.v \end{vmatrix} = K.\begin{vmatrix} x_\rho u_\rho & x_\rho v_\rho \\ y_\sigma u_\sigma & y_\sigma v_\sigma \end{vmatrix};$$

to prove this consider the expression

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} . \begin{vmatrix} x_1 & x_2 \\ y_1 & y_2 \end{vmatrix} . \begin{vmatrix} u_1 & u_2 \\ v_1 & v_2 \end{vmatrix};$$

multiplying by the law of multiplication of determinants the second and the third factors and writing K for the first, we get 18.6; applying the law of multiplication of determinants first to the first two factors, then to the resulting determinant and the third factor we get 18.5; we may thus write

18.7 $$\begin{vmatrix} s(x,u) & s(x,v) \\ s(y,u) & s(y,v) \end{vmatrix} = K.\begin{vmatrix} x.u & x.v \\ y.u & y.v \end{vmatrix};$$

we may now obtain K without using any system of coordinates by dividing the left hand side by the second factor on the right; the vectors x,y,u,v, may be any arbitrary four vectors, only such that the second factor on the right does not vanish. Setting x = i, u = i, y = j, v = j, where i,j are two coordinate vectors we get formula 18.4; now it is seen to hold for any system of coordinate vectors, so that incidentally we have a new proof of the invariance of K.

Before we leave the topic of ordinary surfaces we want to establish a relation between curvature of surfaces and curvature of curves. The points common to our surface and the xz-plane constitute a plane curve whose equation in the xz-plane may be obtained from $z = f(x,y)$ by setting $y = 0$. It is clear that the x-axis is a tangent to this curve and that the first term of the expansion of z as a function $f(x,0)$ into a power series will be obtained by setting $x_2 = 0$ in 18.2. We have thus

18.11 $$\tfrac{1}{2}a_{11}x_1^2$$

as this first non-vanishing term, and comparing with 18.1 we see that the curvature of the curve is a_{11}, or (18.3) the value $s(i,i)$. We can consider any plane passing through the z-axis as the xz-plane, or any unit vector in the tangent plane as the coordinate vector i; we have thus the result, that the curvature at the point of contact of a curve, resulting from the intersection of the surface with a normal plane is $s(i,i)$, if i is a unit vector common to the tangent plane and the normal plane considered. In other words to every direction in the tangent plane, characterized by a unit vector i corresponds a normal plane containing it, and the curvature of the intersection of that plane with the surface is $s(i,i)$.

We see thus that to every direction in the tangent plane corresponds a definite number $s(i,i)$, the curvature in that direction. As an exercise the reader may try to express the curvature corresponding to a direction in the tangent plane in terms of the angle that direction makes with the x-axis.

As the next step of our discussion whose general aim is to arrive at the most general situation as far as the number of dimensions is concerned both of the space from which we start and the configuration in it that we study, we take a skew curve in ordinary space; first we studied a curve ($n = 1$) in a plane ($N = 2$); then a surface ($n = 2$) in the ordinary space ($N = 3$); now we take up the case $n = 1$, $N = 3$. A curve may be given in general by two equations on the three coordinates x,y,z. Solving these equations for y and z, we represent the curve in the form $y = f(x)$, $z = g(x)$; we again make the assumption that a tangent exists for every point and that for every point, if we take this point as the origin and the tangent as the x-axis, it is possible to solve for y and z, and that the functions f and g can be developed into power series; as a result of the choice of the coordinate system, the two power series will begin with quadratic terms

18.8 $$\tfrac{1}{2}a_2x^2, \qquad \tfrac{1}{2}b_2x^2.$$

If we change the y- and z-axes which fall into the normal plane to the curve, to other axes in the same plane, the form of the developments will not be changed, of course, but the coefficients a_2, b_2 will assume new values; if, however, we consider these coefficients as components of a vector, the vector represented by them will be the same in all coordinate systems. Calling this vector v we may say that the curvature situation of the curve is characterized by the expression

18.9 $$\tfrac{1}{2}v x^2.$$

This expression plays the part of the expressions 18.1 and 18.2 which have occurred in the two preceding situations.

19. Generalizations.

In the preceding section we discussed configurations in the ordinary space, and we could rely on our intuition; everybody can conceive a plane curve, a surface, a twisted curve; we have at our disposal physical objects (drawings, graphs, models) with measurements on which quantities of our theories may be identified successfully. In the investigation we undertake now, we cannot use our intuition any more, and the identifications, when they come, will be of a much less immediate character. We have then to rely on analogy with the configurations studied in the preceding section and on mathematical reasoning supported by formulas.

We begin with what seems the next simplest case, a surface in four-dimensional space; it may be considered as a generalization both of a surface and of a curve in ordinary space. Such one is given, in general, by two equations on the four coordinates; in other words, we define as a surface in four-space the totality of points whose coordinates satisfy two equations $F(x,y,z,t) = 0$, $G(x,y,z,t) = 0$ where F,G are two functions subjected to certain restrictions to be imposed presently. We define a plane in four dimensions as a surface which may be given by two linear equations (this definition, although given in terms of coordinates, is invariant, because it can be proved that if the equations are linear in one system of coordinates they remain linear after a transformation; the equivalence of this definition of the plane with that given in Section 7 is easily recognized). In the general case we choose a point on the surface as the origin of coordinates; we solve the two equations for two of the coordinates, and we define as the tangent plane at that point the plane whose equations result from omitting all but linear terms in the expansions. We next choose that plane as one of our coordinate planes; the lowest terms in the expansions are then the quadratic ones; denoting the coordinates for which the equations appear as solved by x_3, x_4, and the two other

coordinates by x_1, x_2, we may write the groups of quadratic terms in the two expansions as

19.1
$$\tfrac{1}{2}(a_{11}x_1^2 + 2a_{12}x_1x_2 + a_{22}x_2^2),$$
$$\tfrac{1}{2}(b_{11}x_1^2 + 2b_{12}x_1x_2 + b_{22}x_2^2).$$

For every vector in the tangent plane of components x_1, x_2 this gives us two numbers which may be considered as the components of a vector in the normal x_3, x_4 plane; or, we may consider this vector as given by a vector form

19.2 $$v_{11}x_1^2 + 2v_{12}x_1x_2 + v_{22}x_2^2 ,$$

the coefficients v_{ij} being vectors of the normal plane whose components are a_{ij} and b_{ij} (along the x_3 and x_4 axes respectively). This expression assigns to every vector of the tangent plane a vector of the normal plane; we may substitute for it, as we did in an analogous case before (compare 18.3), a more general expression

19.3 $$s(x,y)$$
$$= v_{11}x_1y_1 + \tfrac{1}{2}v_{12}x_1y_2 + v_{21}x_2y_1 + v_{22}x_2y_2,$$

where $$v_{21} = v_{12},$$

but although this is linear in each of the vectors x and y it is not a tensor, because the values of this expression are not numbers (they are vectors in the normal plane). We shall not introduce a special name for such expressions because we shall not have to deal with them much; the expression 19.3 we have denoted, as before, by $s(x,y)$, but we must keep in mind that the values of $s(x,y)$ are not numbers but vectors of the normal plane.

We may in this case form the expression 18.5 where it is understood that in the expansion of the determinant scalar products have to be used where ordinary products were used before; this change is made necessary by the more than once mentioned fact that the values of the elements are vectors. In all other respects we can apply to the expression the same reasoning as before and we come to the conclusion that the relation 18.7 remains true, where K is a number, independent of the coordinate systems in the tangent and normal planes, but which after such coordinate systems have been chosen can be calculated in terms of the coefficients of the vector form 19.3 by means of the formula

19.4 $$K = \begin{vmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{vmatrix} .$$

The important fact is that, although our expression 19.3 is a vector expression, and therefore, does not furnish us a tensor, the invariant corresponding to 19.4 is still a number. The other invariant, which could be called mean curvature and written as $v_{11} + v_{22}$ is not a number in this case. The number, K, we call, as before, the total curvature of the surface at the point considered.

In terms of the coefficients of the numerical forms 19.1 the total curvature K may be expressed as follows: expanding the determinant 19.4 we have $K = v_{11}.v_{22} - v_{21}.v_{12}$; the term $v_{11}.v_{22}$, for instance, is the scalar product of the vectors v_{11} and v_{22} whose components are respectively a_{11}, b_{11} and a_{22}, b_{22}; the scalar product $v_{11}.v_{22}$ is then $a_{11}a_{22} + b_{11}b_{22}$; in the same way, the term $v_{21}.v_{12}$ in the expression for K is $a_{12}a_{21} + b_{12}b_{21}$, so that we have for K in terms of the a's and b's rearranging terms and using determinant notation

19.41 $$K = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} + \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} .$$

The next generalization is an easy one; we still consider a surface (n = 2) but instead of a four-dimensional space we take a space of an arbitrary number of dimensions N; we denote N - n = N - 2 by r and we have, a tangent plane, as before, but instead of a normal plane, we have now a normal r-dimensional space, an r-flat as we may say. We call the corresponding coordinates x_3, x_4, etc., or x_{2+k}, where k goes now from 1 to r instead of only taking the values 1 and 2; we have here a vector form which may be written as before (19.2) only the v's are vectors of the normal flat and have r components each; these components we may distinguish by upper indices in brackets; if we denote by I_k the r coordinate vectors in the normal flat, and denote by $a_{ij}^{(k)}$ the components corresponding to I_k of v_{ij} we may write

19.5 $$v_{ij} = \sum_{k=1}^{r} I_k . a_{ij}^{(k)}$$

and for $s(x,y)$ we may write

19.6
$$\sum_{k=1}^{r} (a_{11}^{(k)}x_1y_1 + a_{12}^{(k)}x_1y_2 + a_{21}^{(k)}x_2y_1 + a_{22}^{(k)}x_2y_2)I_k;$$

otherwise there will be no changes. We can form the expression 19.4 as before; it will be independent of the choice of the coordinates x_{2+k} because the scalar products used in the expansion of the determinant are; substituting the values 19.5 and evaluating we will have

19.42 $$K = \sum_{k=1}^{r} \begin{vmatrix} a_{11}^{(k)} & a_{12}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} \end{vmatrix} ,$$

an obvious generalization of formula 19.41 which

may be obtained from this by taking r = 2, and writing a for $a^{(1)}$ and b for $a^{(2)}$ with proper subscripts. We may, if we wish, write out an expression analogous to 18.5. Substituting for s(x,u) etc., the expressions 19.6 and using scalar products in the evaluation of the determinant we shall find

$$19.43 \qquad \begin{vmatrix} s(x,u) & s(x,v) \\ s(y,u) & s(y,v) \end{vmatrix} = \sum_{k=1}^{r} \begin{vmatrix} a^{(k)}_{\alpha\gamma} x_\alpha u_\gamma & a^{(k)}_{\alpha\delta} x_\alpha v_\delta \\ a^{(k)}_{\beta\gamma} y_\beta u_\gamma & a^{(k)}_{\beta\delta} y_\beta v_\delta \end{vmatrix},$$

where the Greek letter subscripts imply summation from one to two corresponding to the tangent plane, and the summation with respect to k corresponding to the r coordinates of the normal flat is indicated in the usual fashion. Each of the determinants corresponding to the different values of k is exactly of the same nature as 18.5 so that the reasoning which led from it to 18.7 applies to each of the determinants without change, and it is easy to see that formula 18.7 continues to hold. We may use this formula to define K which we continue to call total curvature.

And now we come to the last generalization. We consider, in a space of an arbitrary number of dimensions N a curved space of n dimensions with n also an arbitrary number $< N$, which by definition is the totality of points whose coordinates satisfy N - n = r equations

$$19.7 \qquad F_k(x_1, x_2, \ldots x_N) = 0 \quad (k = 1,2,\ldots r).$$

We assume that for every point in the curved space these equations can be solved for r of the coordinates and that these solutions can be expanded into power series in the remaining n coordinates, converging in the neighborhood of the point selected. By a transformation of coordinates we may arrange it so that these expansions begin with second degree terms so that we may write

$$19.71 \qquad x_{n+k} = \tfrac{1}{2} a^{(k)}_{\alpha\beta} x_\alpha x_\beta + \text{ terms of higher degree}$$

where the summation indicated by the Greek indices now goes from 1 to n. The sub-space defined by the first n coordinate axes we call the tangent flat space at the point considered, and the sub-space corresponding to the remaining r coordinate axes - the normal flat space at that point.

As before, we use the coefficients $a^{(k)}_{ij}$ to form the expressions

$$19.8 \qquad a^{(k)}_{\alpha\beta} x_\alpha y_\beta$$

where x_i, y_i are two vectors of the tangent flat; and we combine these expressions into a vector expression

$$19.9 \qquad s(x,y) = \sum_{k=1}^{r} a^{(k)}_{\alpha\beta} x_\alpha y_\beta \, . I_k \, .$$

It is natural to try to generalize the idea of total curvature. We can form the expression 18.5 but, and this is important, the transformation 18.7 does not apply; it was based essentially on the fact n = 2, and it breaks down here.

20. The Riemann Tensor.

The way out of this difficulty is very simple. Although relation 18.7 does not hold we still may consider its left hand side; it is a function of the four vectors x,y,u,v, function, which has numerical values; it is easy to show that it is linear in each of the vector arguments (we leave this proof to the reader because the result will follow later from formula 20.8); it is therefore a tensor, a tensor of rank four; we call it the Riemann tensor, denote it by R(x,y;u,v) and write

$$20.1 \qquad R(x,y;u,v) = \begin{vmatrix} s(x,u) & s(x,v) \\ s(y,u) & s(y,v) \end{vmatrix} .$$

We have then at every point of the curved space a tensor of rank four instead of a number; it is connected with the second degree terms of the expansions 19.71 and therefore characterizes, at least in part, the deviation of the expressions of the x_{n+k} from linearity, or of our space from flatness. The Riemann tensor tells us then something about the curvature of the curved space, and it is often called the curvature tensor.

The situation we have now reminds us of a situation in Section 18. The curvature of a curve was a number; when we passed to a surface we found that its curvature was characterized by a tensor; we have succeeded to derive from this tensor a number K, so that we could express (at least partially) the curvature of a surface by a number. Now passing to higher curved spaces we again obtain a tensor. In the preceding situation we succeeded in interpreting the tensor s(x,y) given by formula 18.3 in terms of curvatures of certain curves on the surface; we found that the value s(i,i) gives the curvature of the normal section determined by the unit vector i and the normal to the surface. Is it possible to interpret the Riemann tensor in an analogous way as giving the total curvatures of some surfaces on our curved space? This is a natural question to ask, and the answer is affirmative. We shall prove, in fact, that certain values of the Riemann tensor give us total curvatures of surfaces situated on the curved space. Let i,j be two arbitrary mutually perpendicular unit vectors of the tangent flat; choose a set of coordinate vectors so that i,j be two of them. Pass through i,j and the normal flat, i.e., through the r vectors I_k,

a flat space of 2 + r dimensions; its points will be those points of the N-space whose coordinates x_3, x_4,....x_n vanish; the intersection of this flat space with the given curved space will be a surface, i.e., a two-dimensional curved space, because the coordinates of its points must satisfy the r equations of the curved space (19.7) and n - 2 equations

$$20.2 \qquad x_3 = 0, \quad x_4 = 0, \ldots . x_n = 0,$$

which together is N - n + n - 2 = N - 2 equations. This surface we may consider as a surface of the r + 2 dimensional flat space 20.2; its equations in that space will be obtained by setting $x_3 = x_4 = \ldots x_n = 0$ in the equations 19.71 of the curved space (just as the equation of the normal section of a surface in the xz-plane was obtained (preceding formula 18.11) by setting y = 0 in the equation of the surface); these equations will then become

$$20.3 \qquad x_{n+k} = \tfrac{1}{2}(a^{(k)}_{11}x_1^2 + 2a^{(k)}_{12}x_1x_2 + a^{(k)}_{22}x_2^2) + \text{terms of higher degree}$$

and the total curvature of this surface is

$$\begin{vmatrix} v_{11} & v_{12} \\ v_{21} & v_{22} \end{vmatrix}$$

with

$$v_{ij} = \Sigma a^{(k)}_{ij} . I_k;$$

but $v_{11} = s(i,i)$, $v_{12} = s(i,j)$, $v_{21} = s(j,i)$, $v_{22} = s(j,j)$, so that we have

$$K = \begin{vmatrix} s(i,i) & s(i,j) \\ s(j,i) & s(j,j) \end{vmatrix}$$

which is R(i,j;i,j), and our statement is proved. As we saw at the end of Section 18, a unit vector i in the tangent plane to an ordinary surface determines a direction, a straight line, which, together with the normal determines a normal plane, and the intersection of that normal plane with the surface is a normal section of curvature s(i,i); here we have the situation that two unit vectors i,j in the tangent flat to a curved space determine an orientation, a plane, which together with the normal r-flat determines a normal r + 2 flat, and the intersection of that normal r + 2 flat with the curved space is a normal section, a surface of curvature R(i,j;i,j). We see then that the Riemann tensor plays with respect to a curved space a role analogous to that played by the tensor s(x,y) with respect to an ordinary surface; our expectations then are fulfilled; we need, it is true, for the purposes of identification with the complete tensor T_{ij} a tensor of the second rank, but we shall get one of rank two from R(x,y;u,v) by applying to it the operation of contraction.

In the meantime let us study the Riemann tensor, or the curvature tensor as it is called sometimes, as we have it. The Riemann tensor is not a general tensor of rank four. It satisfies the relations

$$20.41 \qquad R(x,y;u,v) = -R(y,x;u,v) = -R(x,y;v,u),$$

$$20.42 \qquad R(x,y;u,v) = R(u,v;x,y),$$

$$20.43 \qquad R(x,y;u,v) + R(x,u;v,y) + R(x,v;y,u) = 0,$$

which are easily verified by using 20.1. The first of these relations says that R is antisymmetric in each of the pairs of the vector arguments, and the second, that it is symmetric in the two pairs.

If we introduce a coordinate system in the tangent flat, by picking four coordinate vectors i,j,k,l or i_1, i_2, i_3, i_4, we may represent the vector arguments (as in Section 9) in the form $x = i_\alpha x_\alpha$, etc., substitute these expressions into R(x,y;u,v) and, by using linearity as defined in 9.6, write the Riemann tensor as

$$20.5 \qquad R_{\alpha\beta;\gamma\delta}\, x_\alpha y_\beta u_\gamma v_\delta,$$

where

$$20.6 \qquad R_{ab;cd} = R(i_a, i_b; i_c, i_d).$$

(We use here the first letters of the alphabet as subscripts, instead of i, etc., as before, in order to avoid confusion with the coordinate vectors which we denote by i.) These are the components of the Riemann tensor in the coordinate system chosen. The relations 20.4 can be written in components as

$$20.71 \qquad R_{ab;cd} = -R_{ba;cd} = -R_{ab;dc}$$

$$20.72 \qquad R_{ab;cd} = R_{cd;ab}$$

$$20.73 \qquad R_{ab;cd} + R_{ac;db} + R_{ad;bc} = 0.$$

Exercise. Prove that the number of independent components of a Riemann tensor for four dimensions is 20.

The vectors of the flat spaces tangent to the curved space may be considered as belonging to the curved space, they may be characterized in terms of the space itself, for instance, by giving direction and length; they are accessible as we may say, to beings who live in the space and for whom points outside the space do not exist. Normal vectors, the function s(x,y) etc., are, on the contrary, not accessible to the inhabitants of the space. We shall confine ourselves for the most to the consideration of these internal properties, properties accessible to the inhabitants; but later in the course of our investigation we shall have to use the

expression of the Riemann tensor in terms of the coefficients $a_{ij}^{(k)}$ and we shall conclude this section by deducing it.

Substituting the expression 19.9 for s into 20.1 and using 20.5 for the left hand side, we find

$$20.8 \qquad R_{\alpha\beta;\gamma\delta}x_\alpha y_\beta u_\gamma v_\delta = \begin{vmatrix} \Sigma a^{(k)}_{\alpha\gamma}x_\alpha u_\gamma \, I_k & \Sigma a^{(k)}_{\alpha\delta}x_\alpha v_\delta \, I_k \\ \Sigma a^{(k)}_{\beta\gamma}y_\beta u_\gamma \, I_k & \Sigma a^{(k)}_{\beta\delta}y_\beta v_\delta \, I_k \end{vmatrix};$$

this determinant may be presented as the sum of r^2 determinants, of which, however, only r are different from zero, namely those in which the same I appears in the two columns, because in the expansions of the others all terms vanish as involving products of different and therefore mutually perpendicular I's. What remains is (compare 19.43)

$$\sum_{k=1}^{r} \begin{vmatrix} a^{(k)}_{\alpha\gamma}x_\alpha u_\gamma \, I_k & a^{(k)}_{\alpha\delta}x_\alpha v_\delta \, I_k \\ a^{(k)}_{\beta\gamma}y_\beta u_\delta \, I_k & a^{(k)}_{\beta\delta}y_\beta v_\delta \, I_k \end{vmatrix}$$

$$= \sum_{k=1}^{r} \begin{vmatrix} a^{(k)}_{\alpha\gamma}x_\alpha u_\gamma & a^{(k)}_{\alpha\delta}x_\alpha v_\delta \\ a^{(k)}_{\beta\gamma}y_\beta u_\gamma & a^{(k)}_{\beta\delta}y_\beta v_\delta \end{vmatrix} \cdot$$

because the I's are unit vectors and $I_k . I_k = 1$, or

$$\sum_{k=1}^{r} \begin{vmatrix} a^{(k)}_{\alpha\gamma} & a^{(k)}_{\alpha\delta} \\ a^{(k)}_{\beta\gamma} & a^{(k)}_{\beta\delta} \end{vmatrix} x_\alpha y_\beta u_\gamma v_\delta .$$

Comparing this to the left hand side of 20.8 we have the required expression

$$20.9 \qquad R_{ab;cd} = \sum_{k=1}^{r} \begin{vmatrix} a^{(k)}_{ac} & a^{(k)}_{ad} \\ a^{(k)}_{bc} & a^{(k)}_{bd} \end{vmatrix} .$$

21. Vectors in General Coordinates.

In the last section we learned how to associate with every point of a curved space a tensor of rank four; for our physical interpretation we need one of rank two; but we know how to obtain from a tensor of rank four one of rank two; we have to apply the operation of contraction. The result we shall call the "contracted Riemann tensor" and we shall expect to identify it with the tensor T. The first question we have to ask ourselves in this connection is whether the contracted Riemann tensor satisfies the equation 11.4, viz., $\partial T_{i\alpha}/\partial x_\alpha = 0$. But before we do that we have to go through quite a lengthy development because at the present stage we do not know how to differentiate tensors on a curved space. In flat space we could consider the differential of a vector, or, more exactly, of a vector field, by (roughly) considering the difference of two vectors of the field in two neighboring points. In curved space, or on a curved surface two vectors in two different points belong to two different tangent planes and their difference is not a vector of the surface at all. Or we could in a flat space adopt a cartesian system of coordinates and introduce as the components of the differential the derivatives of the components of the given tensor. This method also is not applicable directly to curved space because there is no such a thing here as cartesian coordinates. Each method could be so modified as to apply to curved space - the geometrical method and the coordinate method. We shall develop here the coordinate method because in addition to permitting us to introduce differentiation - our immediate concern now - it is indispensable in treating special cases.

As we said before, there is no such a thing as the cartesian system of coordinates in curved space, because there are no straight lines; so we shall have to use some other coordinates, let us say, general coordinates; the main difficulty in treating curved spaces is just this, viz., that rectilinear coordinates are not applicable here, or we may say: part of the difficulty lies in the fact that we have to use curvilinear coordinates (the greater part) and part - in the fact that the situation itself is so different from that we encounter in flat space and with which we are more or less familiar. Or to put it in a still different form: the difficulty is two-fold, we have a new material to work on and we have to use new tools. To obviate the difficulty we are going to divide it; we already have studied curved spaces in the preceding sections; now we shall try to become familiar with the new tool - the method of general coordinates, applying it to the old material - ordinary three-dimensional space; and then - beginning with Section 25, we shall study curved space by means of curvilinear coordinates.

The essential thing in the matter of coordinates is that points receive <u>names</u>, the names being composed of numbers, so that we can handle numbers, which we can do by means of formulas, instead of points themselves. There are many different ways of establishing a correspondence between points and triples of numbers; in the one that bears the name of Descartes (Cartesius) the three numbers which are assigned to a point are its distances from three mutually perpendicular planes; there does not seem to be anything that can take the place of this method in a general curved space because there are no planes and straight lines; still we may use coordinates; a system of coordinates on a special curved surface is known to everybody, even to those who never studied analytic geometry; we mean the system of specifying the position on the surface of the earth by means of latitude and longitude. Polar coordinates in the plane or in space furnish another example of a coordinate system which is not cartesian; in what follows we shall

use an entirely arbitrary system of coordinates; we shall assume that a one-to-one correspondence is established between the points of a certain portion of space (which may be the whole space) and the triples of a certain set of triples of numbers. We shall call these numbers u_1, u_2, u_3 or u_i and we shall keep the notation x_i for some definite system of cartesian coordinates. To every triple u_1, u_2, u_3 corresponds a point whose cartesian coordinates (in some definite system) are x_i; these numbers x_i are therefore determined by the u_i's; we have three functions

$$21.1 \qquad x_1 = x_1(u_1,u_2,u_3) \quad x_2 = x_2(u_1,u_2,u_3) \quad x_3 = x_3(u_1,u_2,u_3)$$

which are defined on a certain range of triples. Conversely, if x_1, x_2, x_3 are the cartesian coordinates of a point of the portion of our space for which general coordinate have been introduced, they determine three numbers u_1, u_2, u_3, which therefore are functions of the x's

$$21.2 \qquad u_1 = u_1(x_1,x_2,x_3), \quad u_2 = u_2(x_1,x_2,x_3), \quad u_3 = u_3(x_1,x_2,x_3),$$

which are defined for a certain range of triples x_i and are the inverse functions of the functions 21.1.

We have to handle vectors even more often than we have to handle points, and we want to have a numerical representation also for vectors. Together with cartesian coordinates for points goes a very simple numerical representation for vectors; we represent a vector by three numbers which are the differences between the corresponding coordinates of its end-points, and are called the components of the vector; of course, in a different system of cartesian coordinates the same vector will have other components, but as long as we keep to a definite coordinate system, vectors, as well as points have definite names. The method of representing vectors by their cartesian components has the great advantage that two equal vectors have equal components, that we can add vectors by adding their components, and multiply a vector by a number by multiplying the components by that number; these advantages are peculiar to the cartesian method and cannot be reproduced in other systems. The theory of curved space is <u>differential geometry</u>, we cannot handle immediately by its methods such things as a configuration consisting of two points at a finite distance; if we do we have to introduce intermediate points, instead of subtraction we have here differentiation. There are two ways in which vectors arise by differentiation, and each gives rise to a system of notation for vectors associated with a given coordinate system for points - only for the rectangular cartesian system do the two representations coincide. The two ways in which a vector appears as a result of differentiation are - the tangent vector of a curve and the gradient of a field. In this and the next section we shall take only the first of these two points of view. Given a curve in cartesian coordinates in parametric form

$$21.3 \qquad x(p), \quad y(p), \quad z(p),$$

the components of the tangent vector are obtained by differentiation

$$21.4 \qquad dx/dp, \quad dy/dp, \quad dz/dp.$$

This vector is determined not by the curve alone, but by the particular parametric representation we are using, but in this chapter we are not going to change the parameter often and we shall speak of a curve when we mean "curve in a given parametric representation", and of the tangent vector, when we mean "tangent vector resulting from differentiation with respect to that particular parameter". In cartesian coordinates, then, the components of the tangent vector are obtained by differentiating the coordinates of the points of the curve. This is certainly convenient, and we may ask ourselves whether we could not reproduce this advantage in general coordinates. Let us try; the parametric representation of the curve in the u's can be obtained by substituting $x_i(p)$ into 21.2; the u_i's become then functions of p, and this gives a parametric representation of the same curve in general coordinates; let us <u>agree</u> to represent the vector which, when we used the cartesian system had the components 21.4, by the three numbers

$$21.5 \qquad du_1/dp, \quad du_2/dp, \quad du_3/dp.$$

We have then the required system of representation; but it is not necessary, every time when we want to represent a vector to introduce a curve to which it is tangent; we shall show how to find the components 21.5 when we are given the cartesian components 21.4 without actually considering the curve.

We have, considering that u_1 depends on x_1, x_2, x_3 which in turn depend on p,

$$du_1/dp = \partial u_1/\partial x_1 . dx_1/dp + \partial u_1/\partial x_2 . dx_2/dp + \partial u_1/\partial x_3 . dx_3/dp$$

or using the summation convention and applying to u_1, u_2, u_3,

$$21.6 \qquad du_i/dp = \partial u_i/\partial x_\alpha . dx_\alpha/dp,$$

so that, if we denote the quantities 21.4 by v_i, and the quantities 21.5 by V^i, we have

$$21.7 \qquad V^i = \partial u_i/\partial x_\alpha . v_\alpha .$$

Introducing the abbreviation

21.8 $$a_{ij} = \partial u_i/\partial x_j$$

we may write the last formulas as

21.9 $$V^i = a_{i\alpha}v_\alpha.$$

It will be explained later (Section 23) why we use in the left hand side the index as a superscript. These are transformation formulas for vector components which are associated with the transformation formulas 21.2 for the coordinates of points; the formula just written out permits to find the general components when the cartesian components are given. In a similar way we can find the inverse transformation formulas by starting with a parametric representation of a curve in general coordinates, substituting the $u_i(p)$ into the formulas 21.1 and differentiating; we arrive thus at

21.10 $$v_i = b_{i\alpha}V^\alpha,$$

where

21.11 $$b_{ij} = \partial x_i/\partial u_j.$$

Before we go further we shall use the fact that the formulas 21.9 and 21.10 are inverses of each other to obtain some relations on the a's and b's. Substituting 21.9 into 21.10 we get

21.12 $$v_i = b_{i\alpha}a_{\alpha\beta}v_\beta;$$

the left hand side may be written as $\delta_{i\beta}v_\beta$ so that

$$b_{i\alpha}a_{\alpha\beta}v_\beta = \delta_{i\beta}v_\beta,$$

and since this is an identity (v being arbitrary) we have

21.13 $$b_{i\alpha}a_{\alpha j} = \delta_{ij}.$$

In the same way we may obtain

21.14 $$a_{i\alpha}b_{\alpha j} = \delta_{ij}.$$

We want to be able to operate with general components of vectors, for instance, find a scalar product of two vectors given in their general components; it is easy to obtain a formula answering this question by passing to cartesian components, and then applying the formula for the scalar product in cartesian components. Let the general components of two vectors be V_i and W_i; according to the formula 21.10 their cartesian components are

$$v_i = b_{i\alpha}V^\alpha \quad \text{and} \quad b_{i\beta}W^\beta = w_i,$$

where we use another summation letter in the second expression to avoid complications in what follows. Now we write the scalar product using the formula $v_\gamma w_\gamma$ and get

21.15 $$b_{\gamma\alpha}V^\alpha b_{\gamma\beta}W^\beta = b_{\gamma\alpha}b_{\gamma\beta}V^\alpha W^\beta = g_{\alpha\beta}V^\alpha W^\beta.$$

It follows that in order to be able to find scalar products of vectors given by their general components we have to know the quantities

21.16 $$g_{ij} = b_{\gamma i}b_{\gamma j}.$$

The quantities a's and b's help to pass from a certain cartesian system of coordinates to the general system; they express a relation between the general system and that particular cartesian system; and thus are not of fundamental importance; the quantities g_{ij}, on the contrary, although they have been obtained by means of the a's and b's, are independent, as their significance shows, from any particular system of cartesian coordinates, they characterize the system of coordinates we are using in itself (and, as we shall see later, they characterize it completely, so that the g's are all we need to know in order to be able to handle vectors given by their general components). The a's as well as the b's may be considered either as functions of the x's, or as functions of the u's. The g's always will be considered as functions of the u's.

Before we go any farther we note that, as it immediately follows from the definition 21.16 the g's are symmetric in the indices:

21.16' $$g_{ij} = g_{ji}.$$

In order to show the importance of the g's let us deduce a formula for the length of a curve given in general coordinates. Let the curve be given by

21.31 $$u_i(p).$$

For a curve given in cartesian coordinates we assume as known the formula (compare 12.1)

21.17 $$s = \int \sqrt{(dx/dp)^2 + (dy/dp)^2 + (dz/dp)^2}\,dp$$

where s is the arc length between two points. This formula involves three inverse operations, that of integration, that of taking the square root, and that of division. It is not pleasant, in general, to have to do with these operations, and so we shall free our formula from them, and write it as

21.17' $$ds^2 = dx^2 + dy^2 + dz^2.$$

The formula as just written is not essentially different from the one written above, and means exactly the same thing. The sign d may be taken

to mean differentiation with respect to some unspecified parameter, since the correctness of the formula does not depend on what parameter we are using (provided that the same parameter is used on both sides). We translate 21.17' now into general components. Differentiating 21.3 we have

$$dx = b_{1\alpha}.du^\alpha, \quad dy = b_{2\alpha}.du^\alpha, \quad dz = b_{3\alpha}.du^\alpha;$$

for dx^2 we may write $b_{1\alpha}du^\alpha.b_{1\beta}du^\beta$; using similar expressions for the other terms of 21.17' we get

$$\begin{aligned} ds^2 &= (b_{1\alpha}b_{1\beta} + b_{2\alpha}b_{2\beta} + b_{3\alpha}b_{3\beta})du_\alpha du_\beta \\ &= g_{\alpha\beta}du_\alpha du_\beta, \end{aligned} \qquad 21.18$$

using the abbreviation 21.15 introduced before.

We see that the quantities g_{ij} appear again.

22. Tensors in General Coordinates.

We come now to the representation of tensors. We know that a tensor is a function which assigns numbers to vectors, and the question of representation will be simply this: given the general components of the vector arguments how to find the corresponding value of the tensor. We have already solved this problem for one particular tensor, namely for the tensor of the scalar product which we expressed in the preceding section by means of the g's, and we shall use the same method in the general case. Given the cartesian components of a tensor f_{ij} and the general components V^i and W^i of two vector arguments, to find the corresponding value of the tensor. We pass from general components to cartesian components by formulas 21.10 and substitute these expressions into the expression $f_{\gamma\delta}v_\gamma w_\delta$ for the value of the tensor; the result is

$$f_{\gamma\delta}b_{\gamma\alpha}V^\alpha b_{\delta\beta}W^\beta = f_{\gamma\delta}b_{\gamma\alpha}b_{\delta\beta}V^\alpha W^\beta,$$

and this may be written as

$$F_{\alpha\beta}V^\alpha W^\beta = f_{\gamma\delta}v_\gamma w_\delta \qquad 22.1$$

if we set

$$f_{\gamma\delta}b_{\gamma i}b_{\delta j} = F_{ij}\ ; \qquad 22.2$$

we call F_{ij} the general components of the tensor, whose cartesian components are f_{ij}, and we see that the values of a tensor are expressed by formula 22.1 in its general components and the general components of the vector arguments in the same way as in terms of cartesian components of the tensor and the vector arguments. Formula 21.15 for the scalar product of two vectors may be considered as a special case of 22.1. The cartesian components of the tensor g_{ij} are, of course, the δ_{ij}, and substituting δ for f in 22.2 we get the expressions 21.16 for F_{ij}. We treated here as an example a tensor of rank two; similar calculations can be performed for a tensor of any rank; we give the results for rank one and three, leaving it to the reader to go through the calculations:

$$F_i = f_\gamma b_{\gamma i}, \quad F_{ijk} = f_{\alpha\beta\gamma}b_{\alpha i}b_{\beta j}b_{\gamma k}. \qquad 22.11$$

Now naturally the inverse problem presents itself; given the general components of a tensor to find its cartesian components. The problem can be solved by substituting in the expression for the value of a tensor (we again take a tensor of rank two as an example) given in terms of general components, $F_{\gamma\delta}V^\gamma W^\delta$, the expressions 21.9 for the general components of the vector arguments in terms of the cartesian components: $V^i = a_{i\alpha}v_\alpha$, $W^i = a_{i\beta}w_\beta$; the result is

$$F_{\gamma\delta}a_{\gamma\alpha}v_\alpha a_{\delta\beta}w_\beta = F_{\gamma\delta}a_{\gamma\alpha}a_{\delta\beta}v_\alpha w_\beta;$$

comparing this to the expression $f_{\alpha\beta}v_\alpha w_\beta$ for the value of a tensor in cartesian components we derive the desired transformation formula for passing from general to cartesian components; here are these formulas for the first three ranks

$$\begin{aligned} f_i &= F_\alpha a_{\alpha i}, \qquad f_{ij} = F_{\alpha\beta}a_{\alpha i}a_{\beta j}, \\ f_{ijk} &= F_{\alpha\beta\gamma}a_{\alpha i}a_{\beta j}a_{\gamma k}. \end{aligned} \qquad 22.3$$

We know now how to write tensors in general components, and we want to find out how to perform operations on them. Of course, we could always pass from general components to cartesian components, perform the required operations and then, if the result is a tensor, pass back to general components; but instead of following this program in every special case as it presents itself we shall do it once for all and derive general formulas whose application in special cases is much more convenient than ad hoc calculations.

We begin with the operation of contraction. Given again a tensor of rank two by its general components F_{ij} we pass to its cartesian components by formula 22.3 and now we contract by taking the sum of components with equal indices according to the original definition,

$$f_{\gamma\gamma} = F_{\alpha\beta}a_{\alpha\gamma}a_{\beta\gamma} = F_{\alpha\beta}g^{\alpha\beta} \qquad 22.4$$

where we use the abbreviation

$$a_{i\gamma}a_{j\gamma} = g^{ij}. \qquad 22.5$$

For tensors of higher ranks (contraction is possible only for tensors whose rank is ≥ 2) entirely analogous formulas may be obtained easily; indices which are not affected by contraction may be simply disregarded, as it follows from similar calculations which are left to the

student. For example, the result of contracting with respect to the second and fourth indices of a tensor of the fourth rank F_{ijkl} will be

22.41 $$g^{\alpha\beta}F_{i\alpha k\beta}.$$

The quantities g^{ij} introduced a moment ago play quite an important role comparable to that of the g_{ij} with lower indices, and they are connected with them by the formula

22.6 $$g^{i\alpha}g_{\alpha j} = \delta_{ij}.$$

To prove it suffices to substitute the expressions 21.16 and 22.5 for the two kinds of g's and to apply formula 21.14 twice; we may also notice that, as it follows from the definition,

22.7 $$g^{ij} = g^{ji},$$

so that formula 22.6 may also be written as

22.61 $$g_{i\alpha}g^{\alpha j} = \delta_{ij}.$$

Next comes the operation of differentiation. The result of differentiating a tensor is always a tensor of rank higher by one than the given tensor; its components will have one more index than those of the given tensor; we shall denote them by simply adding a new index preceded by a comma, to the symbol of the given tensor.

Because the situation is slightly more complicated, let us start in translating differentiation into general components with the simplest case of a tensor of rank <u>one</u> given in general components, F_i. We follow the same program: as a first step we pass to cartesian components by formula 22.3 and get $f_i = F_\alpha a_{\alpha i}$: we next find the cartesian components of the differential by simply differentiating with respect to cartesian coordinates with the result

$$f_{i,j} = \frac{\partial}{\partial x_j}f_i = \frac{\partial}{\partial x_j}(F_\alpha a_{\alpha i}).$$

As a third step we pass back to general components using the formula 22.2 and arriving at the result

$$F_{i,j} = f_{\gamma,\delta}b_{\gamma i}b_{\delta j} = \frac{\partial}{\partial x_\delta}(F_\alpha a_{\alpha\gamma})b_{\gamma i}b_{\delta j};$$

but $b_{\delta j}$ according to formula 21.11 is $\partial x_\delta/\partial u_j$, so that this expression reduces to

$$\frac{\partial}{\partial u_j}(F_\alpha a_{\alpha\gamma})b_{\gamma i} = a_{\alpha\gamma}b_{\gamma i}\frac{\partial}{\partial u_j}F_\alpha + F_\alpha b_{\gamma i}\frac{\partial}{\partial u_j}a_{\alpha\gamma};$$

using relation 21.14 the first term reduces to

$$\partial F_i/\partial u_j,$$

just what we would expect from analogy with cartesian coordinates as an expression for the result of differentiation; however, this is not the whole answer because there is a second term so that the final result is

22.8 $$F_{i,j} = \partial F_i/\partial u_j - \Gamma^\alpha_{ij}F_\alpha$$

where we set

22.81 $$\Gamma^k_{ij} = -b_{\gamma i}\frac{\partial}{\partial u_j}a_{k\gamma};$$

the second term may be considered as being in the nature of a correction to the expected result; we call it the correction term, and we call Γ^k_{ij} the correction coefficients. We see then that in general coordinates the components of the differential of a tensor of rank one consist of two parts - the first expresses the change (or rate of change) of components of the tensor, the second is due to the change of the coordinate system from point to point. In the case of the cartesian system the coordinate system is, so to say, the same in all points, the second term is zero (the a's reduce in this case to constants, and their derivatives vanish); another extreme example is furnished by a tensor whose components are constants in some non-cartesian system of coordinates (for instance, polar coordinates); the derivatives of the components with respect to the coordinates are zero but the components of the differential are not; their values are given by the correction terms alone.

For tensors of rank higher than the first the calculations are slightly more complicated, but the principle is the same; we write out the results for tensors of rank two and three

22.82 $$F_{ij,k} = \frac{\partial}{\partial u_k}F_{ij} - \Gamma^\alpha_{ik}F_{\alpha j} - \Gamma^\alpha_{jk}F_{i\alpha}$$

22.83 $$F_{ijk,m} = \frac{\partial}{\partial u_m}F_{ijk} - \Gamma^\alpha_{im}F_{\alpha jk} - \Gamma^\alpha_{jm}F_{i\alpha k} - \Gamma^\alpha_{km}F_{ij\alpha};$$

the general rule ought to be clear now; there are as many correction terms as there are indices; each correction term corresponds to one index, the other indices being disregarded in its formation.

In order to be able to perform the operation of contraction (and the operation of scalar multiplication is a special case of it) in general coordinates we have to know, as we saw, the values of the g's; in order to be able to perform the operation of differentiation we have to know the value of the Γ's (the correction coefficients); if we know those we can perform all the necessary operations in general coordinates without going back to cartesian coordinates. We shall show now the values of the Γ's can be derived from those of the g's.

The correction coefficients were given originally by the formula 22.81; we can give to this another form by using the relation 21.14; writing it as $a_{k\gamma}b_{\gamma i} = \delta_{ki}$ and differentiating it with respect to u_j we get, since the δ's are constants,

$$a_{k\gamma}\frac{\partial}{\partial u_j}b_{\gamma i} + b_{\gamma i}\frac{\partial}{\partial u_j}a_{k\gamma} = 0$$

so that we have

$$\Gamma^k_{ij} = a_{k\gamma}\frac{\partial}{\partial u_j}b_{\gamma i}$$

or, recalling the definition 21.11 of the b's,

22.84 $$\Gamma^k_{ij} = a_{k\gamma}\frac{\partial^2 x_\gamma}{\partial u_i \partial u_j}$$

from this expression it follows that Γ is not affected by interchanging the two lower indices, or

22.71 $$\Gamma^k_{ij} = \Gamma^k_{ji}.$$

We may now show how the Γ's can be derived from the g's. We shall do that by using the following artifice. Consider the tensor of scalar multiplication, whose cartesian components are the δ_{ij} and whose general components were shown to be g_{ij}; the components of the differential of this tensor in cartesian components are the derivatives of the δ's and therefore zero; the second formula 22.11 shows that the general components of this tensor of the third rank also must vanish, so that

22.85 $$g_{ij,k} = 0$$

(we did not promise that general components of tensors will always be given by capital letters, but since heretofore we have been using capital letters for them it may be well to emphasize that the g's are intended to represent (following the generally accepted custom) general components of the scalar multiplication tensor). On the other hand, we can calculate the components of $g_{ij,k}$ by the application of formula 22.82 and so we get

$$\frac{\partial g_{ij}}{\partial u_k} - \Gamma^\alpha_{ik}g_{\alpha j} - \Gamma^\alpha_{jk}g_{i\alpha} = 0;$$

this is a system of equations connecting the Γ's with the g's and their derivatives; we want to solve them for the Γ's. For that purpose we write out the above relation in two more forms resulting from it by cyclic interchanges of indices:

$$\frac{\partial g_{jk}}{\partial u_i} - \Gamma^\alpha_{ji}g_{\alpha k} - \Gamma^\alpha_{ki}g_{j\alpha} = 0,$$

$$\frac{\partial g_{ki}}{\partial u_j} - \Gamma^\alpha_{kj}g_{\alpha i} - \Gamma^\alpha_{ij}g_{k\alpha} = 0;$$

subtracting the last two relations from the first we notice that as the result of symmetry of the g's and the Γ's in the lower indices (formulas 21.16' and 22.71) four of the terms containing the Γ's cancel and the remaining two are identical; we thus have

22.9 $$\tfrac{1}{2}\left(\frac{\partial g_{jk}}{\partial u_i} + \frac{\partial g_{ki}}{\partial u_j} - \frac{\partial g_{ij}}{\partial u_k}\right) = \Gamma^\alpha_{ij}g_{k\alpha} = \Gamma_{k,ij}.$$

We multiply now both sides by g^{mk} and sum with respect to k, writing for it a Greek index, e.g., β. Taking into account 22.61 we have

22.91 $$\tfrac{1}{2}g^{m\beta}\left(\frac{\partial g_{j\beta}}{\partial u_i} + \frac{\partial g_{\beta i}}{\partial u_j} - \frac{\partial g_{ij}}{\partial u_\beta}\right) = \Gamma^m_{ij}.$$

This shows how, given the g's, to calculate the Γ's. We see thus that if only we are given the g's as functions of the u's we can perform all the required operations on tensors. Very often the calculation of the Γ's is divided into two parts; first the left hand sides of 22.9 are calculated and listed; they are denoted by $\Gamma_{k,ij}$; and then the Γ^m_{ij} are calculated using the formulas in the form

22.92 $$\Gamma^m_{ij} = g^{m\beta}\Gamma_{\beta,ij}.$$

23. Covariant and Contravariant Components.

We know (Section 9) that a vector is a tensor of rank one, or, more precisely, that to every vector v there corresponds a tensor of rank one v.x which has the same cartesian components. Now we have introduced general components for vectors and also for tensors; if we have cartesian components of a vector v_i to them correspond (21.9) the general components

$$V^i = a_{i\alpha}v_\alpha;$$

also if we consider the v_i as the components of a tensor to them correspond (22.11) the general components

$$V_i = v_\alpha b_{\alpha i};$$

to the same cartesian components v_i there correspond thus two different sets of general components depending on whether we consider the v_i as vector or as tensor components; it was in anticipation of this situation that we have been using the index for general vector components as a superscript. Essentially, a vector and a tensor of rank one are one and the same thing; and so we have two different systems of components for every vector (in a given general coordinate system); the components with subscripts are called covariant components, those with the superscript - contravariant components. It is clear from what precedes, but it may be worthwhile to repeat that we have here two different representations of one and the same thing.

It was mentioned in Section 21 that there are two ways in which a vector results from differentiation; one, a vector considered as a tangent vector to a curve, was discussed before, and is the basis of what we have been doing all this time; it is interesting to consider now briefly the other. If we start with a scalar field $f = f(x,y,z)$ we may derive a vector field by differentiation, and the cartesian components

of this vector field will be

23.1 $$\partial f/\partial x_i;$$

this vector is known as the gradient of the field f; now, we may give the same scalar field in general coordinates

$$f = f\left\{x_1(u_1u_2u_3),\ x_2(u_1u_2u_3),\ x_3(u_1u_2u_3)\right\}$$
$$= F(u_1u_2u_3);$$

if we differentiate f with respect to u_i, will we obtain general components of the gradient vector field? The question is easily answered by computing these components; we have

23.2 $$\frac{\partial f}{\partial u_i} = \frac{\partial f}{\partial x_\alpha}\cdot\frac{\partial x_\alpha}{\partial u_i} = \frac{\partial f}{\partial x_\alpha}\cdot b_{\alpha i};$$

comparing this with 22.11 we see that the partial derivatives $\partial f/\partial u_i$ are the components with subscripts - the <u>covariant</u> components in general coordinates of the gradient. The two representations, the covariant, and the contravariant, may be thus considered as corresponding to two ways in which a vector can be arrived at by differentiation; if we consider a vector as a gradient we arrive naturally at its representation by covariant components, if we consider it as a tangent vector we arrive at its representation by contravariant components. (The name covariant, by the way, is intended to indicate that these components change in the same way, as, or have similar formulas of transformation with, partial derivatives.) In the case of cartesian coordinates, of course, covariant and contravariant components of a given vector coincide: in this case it is not necessary to make any distinction.

We shall have to use covariant as well as contravariant components, and it is important to be able to pass from one to the other representation; the necessary formulas can be found, of course, by passing through a cartesian representation. Let covariant components F_i of a tensor of rank one (or vector) be given; formulas 22.3 show us that the corresponding cartesian components are $f_i = F_\alpha a_{\alpha i}$; in terms of these the contravariant components are obtained by formula 21.9 which gives here

23.3 $$F^i = a_{i\beta}F_\alpha a_{\alpha\beta} = g^{i\alpha}F_\alpha$$

if abbreviation 22.5 is used. In the same way it is easy to prove the following formula, which permits to calculate covariant components when the contravariant components are given

23.31 $$F_i = g_{i\alpha}F^\alpha.$$

It may seem that there is a wasteful redundancy in this double system of notations, that one representation is enough, and that to have two, means to indulge in luxury; as a matter of fact this double notation <u>is</u> a defect from a didactical point of view: it makes it more difficult to learn the new language; but once mastered it makes the calculations much simpler and the formulas much shorter and more elegant, if properly used; as an example, we want to give the formula for the scalar product of two vectors, one of which is given in covariant, and the other in contravariant components. This formula can be obtained by the usual procedure, i.e., passing through cartesian components, but we have already reached a stage where we can dispense to a great extent with the use of cartesian coordinates. The required formula is simply

23.4 $$V^\alpha W_\alpha$$

and it can be proved by simply substituting for W_α its expression according to formula 23.31, viz., $g_{\alpha\beta}W^\beta$ and comparing the result to 21.15; of course, the scalar product could also be written as $V_\alpha W^\alpha$, and also $g^{\alpha\beta}V_\alpha W_\beta$, as it is easy to verify.

In Section 22 we derived a system of representation for tensors starting with the contravariant representation of vectors; we could do the same thing starting with the covariant representation of vectors, and we shall do it so as to have a perfectly symmetrical system of notations.

Suppose we are given covariant components of two vectors V_i, W_i and the components F_{ij} of a tensor, and we want to find the value of the tensor corresponding to the given vectors as arguments; we know how to solve the problem if the vectors are given by their contravariant components; therefore, let us calculate first the contravariant components, viz., $V^i = g^{i\gamma}V_\gamma$, $W^i = g^{i\delta}W_\delta$, and then substitute them into the left hand side of expression 22.1 giving the value of the tensor. The result is

23.4 $$F_{\alpha\beta}V^\alpha W^\beta = F_{\alpha\beta}g^{\alpha\gamma}V_\gamma g^{\beta\delta}W_\delta = g^{\alpha\gamma}g^{\beta\delta}F_{\alpha\beta}V_\gamma W_\delta$$

which may be written as

23.41 $$F^{\gamma\delta}V_\gamma W_\delta$$

if we introduce the notation

23.42 $$F^{ij} = g^{\alpha i}g^{\beta j}F_{\alpha\beta}.$$

We call F^{ij} the <u>contravariant</u> components of the tensor F_{ij} and the components with lower indices (subscripts) which we have been using for tensors heretofore we call <u>covariant</u> components. We have thus two representations not only for tensors of the first rank (vectors) but also for tensors of all ranks. In one case we have been using already a symbol with two superscripts, viz., the g^{ij} (introduced by 22.5); we shall show now that this notation is in agreement

with the general notation we are introducing now by proving that these g's with upper indices are the contravariant components of the tensor of scalar multiplication. In order to prove that, we notice that, according to formula 23.42 the contravariant components of a tensor of covariant components g_{ij} are

$$g^{\alpha i} g^{\beta j} g_{\alpha\beta};$$

but according to 22.61 $g_{\alpha\beta}g^{\beta j} = \delta_{\alpha}^{\ j}$ so that we get $\delta_{\alpha}^{\ j} g^{\alpha i} = g^{ji}$, and the assertion is proved.

Next we want to learn how to differentiate a tensor given in contravariant components, but before we do that it seems necessary to introduce what we call mixed components. Suppose we are given one vector argument of a tensor of rank two in contravariant, and the other in covariant components, and we want to find the value of the tensor; if the components of the two given vectors are v^i and w_i, and the cartesian components of the tensor are f_{ij}, we pass to cartesian components of the vector arguments

$$23.5 \qquad v_i = b_{i\gamma} v^{\gamma}, \qquad w_i = w_{\delta} a_{\delta i},$$

and express the value of the tensor as

$$23.51 \qquad f_{\alpha\beta} v_{\alpha} w_{\beta} = f_{\alpha\beta} b_{\alpha\gamma} a_{\delta\beta} v^{\gamma} w_{\delta} = F_{\gamma}^{\ \delta} v^{\gamma} w_{\delta}$$

where the notation

$$23.52 \qquad F_i^{\ j} = f_{\alpha\beta} b_{\alpha i} a_{j\beta}$$

is introduced. The numbers $F_i^{\ j}$ with one lower and one upper index are called mixed components of the tensor of rank two whose cartesian components are f_{ij}. In this same way we may consider mixed components for a tensor of any rank with as many of the indices up as we may wish, and the others - down.

We can pass from one kind of component to any other directly, without going through cartesian components. The transition from components in which a certain index (for example, the third) is used as a superscript to components in which the same index is used as a subscript is called the lowering of that index. This change does not affect the geometrical meaning of the tensor, it merely corresponds to a transition from an expression of the tensor in which the corresponding vector argument (in our example, the third) was given by its covariant components to an expression of the same tensor using contravariant components for that vector argument. The formula for the lowering of an index is easily found to be independent from all other indices, so that, disregarding them, we always may use 23.31. Formula 23.3 may be considered as a general formula for raising an index. Lowering and raising of indices is sometimes referred to as juggling with indices.

Again it may seem that the introduction of mixed components is superfluous, but there are advantages in using mixed components.

One advantage appears in connection with contraction. The result of contraction is given in terms of covariant components by formula 22.4 (or 22.41). But, according to 23.3 we may write $F^i_{\ j}$ for $F_{\alpha j} g^{i\alpha}$ so that 22.4 may be written as

$$23.43 \qquad F^{\alpha}_{\ \alpha}$$

and for 22.41 we may write $F_i^{\ \alpha}{}_{k\alpha}$. We see thus that if a tensor is given by its mixed components and the two indices with respect to which we contract appear on different levels (one as a subscript, the other as a superscript) contraction is performed (like in cartesian coordinates) by simply replacing each of the two indices by the same Greek letter.

Another case where there is great advantage in using mixed components is that of differentiation of a contravariant tensor (as we say sometimes for: "tensor given by its contravariant components"; a tensor in itself is, of course, neither contravariant nor covariant - covariance and contravariance are only properties, or types of representation of tensors); the components of the differential will have one more index, and this index as one derived by differentiation will naturally be a subscript, whereas the old indices are superscripts; this does not mean that we cannot pull the new index up, or the old ones down, but the expressions resulting from that would be more complicated.

Suppose the given contravariant tensor is of rank one (a vector) v^i; we pass to cartesian components:

$$v_i = b_{i\gamma} v^{\gamma};$$

we differentiate this:

$$v_{i,j} = \frac{\partial v_i}{\partial x_j} = \frac{\partial}{\partial x_j} b_{i\gamma} v^{\gamma},$$

and we pass to mixed components by formula 23.52:

$$v^i_{\ ,j} = a_{i\alpha} b_{\beta j} \frac{\partial}{\partial x_{\beta}} (b_{\alpha\gamma} v^{\gamma}) = a_{i\alpha} \frac{\partial}{\partial u_j} (b_{\alpha\gamma} v^{\gamma})$$

$$= a_{i\alpha} b_{\alpha\gamma} \frac{\partial}{\partial u_j} v^{\gamma} + a_{i\alpha} \frac{\partial b_{\alpha\gamma}}{\partial u_j} v^{\gamma};$$

and this, using 21.14 and 22.71 reduces to

$$23.6 \qquad v^i_{\ ,j} = \frac{\partial v^i}{\partial u_j} + \Gamma^i_{\gamma j} v^{\gamma}.$$

The reader should be able, following the examples given, to deduce formulas for differentiation of a tensor given in any form. We just mention, because we will have occasion to use it later, the formula for differentiation of a mixed tensor

$$23.7 \qquad v^i_{\ j,k} = \frac{\partial v^i_{\ j}}{\partial u_k} + \Gamma^i_{\gamma k} v^{\gamma}{}_j - \Gamma^{\gamma}_{jk} v^i_{\gamma}.$$

We are in possession now of all the formal rules of operations on tensors in general coordinates. Although these rules were deduced by means of cartesian coordinates these coordinates and components together with all formulas involving the a's and the b's form only a kind of scaffolding that can be removed after the building has been completed. All we have to know in order to operate on tensors are the g's. Using the g's we can lower and raise indices and contract and, as a special case, find the scalar product of two vectors; also find the angle between two vectors (using formula 7.5) and the length of a curve (using 21.18). Given the g's we can calculate the Γ's (end of Section 22) and with the aid of the Γ's we can differentiate tensors (formulas 22.8). We see thus that the g's play a fundamental part in all operations - the tensor of which they are components is often called the fundamental tensor.

Before we conclude we might state explicitly that all the formulas we have obtained are entirely independent of the number of dimensions.

24. Physical Coordinates as General Coordinates.

The principal purpose for the introduction of general coordinates was to make possible the treatment of tensors in curved space but it happens that general coordinates may be used with great advantage also in Special Relativity Theory, namely, in connection with the situation arising from the "minus sign". We remedied this situation in Section 4 by introducing imaginary coordinates and tensor components; we know how, using these imaginary quantities to write our formulas in a nice symmetrical form. The system of notations for general coordinates that we have introduced permits us now to reintroduce real quantities, and still to preserve symmetry in the formulas. We shall express our four mathematical coordinates x_1, x_2, x_3, x_4, of which the fourth is imaginary in terms of four real coordinates which we may denote by u_l; we may choose as these four real numbers the physical coordinates x,y,z,t and consider the formulas

$$24.1 \qquad x_1 = x, \quad x_2 = y, \quad x_3 = z, \quad x_4 = it$$

as the transformation formulas, corresponding to 21.1; and

$$24.11 \qquad x = x_1, \quad y = x_2, \quad z = x_3, \quad t = -ix_4$$

as the inverse formulas corresponding to 21.2.

The a_{ij} and the b_{ij} with different subscripts are easily seen to be zero, and we have (compare 21.8 and 21.11)

$$24.2 \qquad \begin{aligned} a_{11} &= a_{22} = a_{33} = 1, \quad a_{44} = i, \\ b_{11} &= b_{22} = b_{33} = 1, \quad b_{44} = -i; \end{aligned}$$

from these we obtain using 21.16

$$24.21 \qquad g_{11} = g_{22} = g_{33} = -g_{44} = 1, \text{ all others zero}$$

and the same values we obtain for the g's with upper indices, using 22.5.

The x,y,z,t may be considered as the contravariant components of the radius vector leading from the origin to the point P; the covariant components of the same vector are seen, applying the formula 23.31, to be x,y,z,-t.

The formula for the square of the distance from the origin may be obtained either from 21.15 or from 23.4; it is (compare 10.1)

$$24.3 \qquad x^2 + y^2 + z^2 - t^2.$$

We come next to Maxwell's equations where the "minus sign trouble" originated. To conform with the notations of this chapter we should use for the cartesian components - the mathematical components of preceding chapters - small letters, so that formulas 4.72 will have to be written

$$f_{41} = iX, \quad f_{42} = iY, \quad f_{43} = iZ, \quad f_{23} = L,$$

$$f_{31} = M, \quad f_{12} = N.$$

Using formulas 22.2 and 24.2 we obtain the covariant components in physical coordinates - and we use here capital letters - as follows:

$$24.4 \qquad \begin{aligned} F_{41} &= X, \quad F_{42} = Y, \quad F_{43} = Z, \\ F_{23} &= L, \quad F_{31} = M, \quad F_{12} = N. \end{aligned}$$

Mixed and contravariant components may be obtained by raising indices - formula 23.2. Due to the simple character of the g's given by 24.21 it is easy to see that raising one of the indices 1,2,3 does not change the numerical value of a component, so that, for instance,

$$24.5 \qquad F^{23} = g^{2\alpha}g^{3\beta}F_{\alpha\beta} = g^{22}g^{33}F_{23} = F_{23},$$

and raising of the index 4 just changes the sign of the component so that

$$24.6 \qquad F^4{}_1 = g^{4\alpha}F_{\alpha 1} = g^{44}F_{41} = -F_{41}.$$

Which components shall we use in Maxwell's equations? It is clear that in the first (11.2) set all the indices must be on the same level, and since the last one must be a lower index we write all of them down. In the second set (11.3) again the one after the comma must be down; but the one with which we contract it must be on the other level, and therefore up; the position of the third index is arbitrary. We have thus, as the Maxwell equations for free space

$$24.7 \qquad F_{ij,k} + F_{ki,j} + F_{jk,i} = 0, \quad F^{i\alpha}{}_{,\alpha} = 0$$

and in the presence of matter the second set becomes

$$24.71 \qquad F^{i\alpha}{}_{,\alpha} = \varepsilon u^i.$$

We notice here that no imaginary quantities appear and in spite of this our formulas are symmetric. The raising of the index 4 is equivalent to changing the sign of a component, and this is how the minus sign is taken care of.

We shall now write out the expressions for the stress energy tensor and the equations of motion; it is clear that the formulas 11.4 and 11.5 become

$$24.8 \qquad T^{i\alpha}{}_{,\alpha} = 0$$

$$24.9 \qquad T^i{}_j = F^i{}_\rho F^\sigma{}_j - \tfrac{1}{4}\delta_{ij} F^\sigma{}_\rho F^\rho{}_\sigma - \rho u^i u_j$$

or

$$24.91 \qquad T^{ij} = F^i{}_\rho F^{\rho j} - \tfrac{1}{4} g^{ij} F^\sigma{}_\rho F^\rho{}_\sigma - \rho u^i u^j$$

or

$$24.92 \qquad T_{ij} = F_{i\rho} F^\rho{}_j - \tfrac{1}{4} g_{ij} F^\sigma{}_\rho F^\rho{}_\sigma - \rho u_i u_j.$$

The continuity equation (11.1) will now be written as

$$24.10 \qquad (\rho u^\alpha)_{,\alpha} = 0.$$

25. Curvilinear Coordinates in Curved Space.

We want next to apply the general coordinate system that we have introduced for flat space also to curved spaces. In flat space we introduce the language of general coordinates and components by translating from the language of cartesian coordinates and components; in curved space we have no cartesian components; we shall have, therefore, to begin by introducing something that will play the role of cartesian coordinates; we shall introduce quasi-cartesian coordinates, which will take that place; but whereas cartesian coordinates are universal in that the same system of coordinates works for the whole plane, or flat space - the neighborhood of every point in curved space has its own system of quasi-cartesian coordinates. They are defined in the following way: Consider at a given point P the tangent flat; there will be in general a neighborhood of P such that no two points of that neighborhood have the same projection on the tangent flat (for a sphere, e.g., we may obtain such a neighborhood by drawing any small circle around the point of contact) for such a neighborhood there exists a one-to-one correspondence between the points of it and the points of the tangent flat which are their projections. We introduce now on the tangent flat a cartesian system of coordinates with origin at the point of contact, and we use the coordinates of a point of the flat as the quasi-cartesian coordinates of the point of the curved space whose projection it is; if, for instance, a surface is given by equation

$$25.1 \qquad z = \tfrac{1}{2}(ax^2 + 2bxy + cy^2) + \text{t.h.d.},$$

x,y will be the quasi-cartesian coordinates of the point x,y,z, of the surface for the neighborhood of 0,0,0; - and in the general case, if the curved space is given by 19.71

$$25.11 \qquad x_{n+k} = \tfrac{1}{2} a_{\alpha\beta} x_\alpha x_\beta + \text{t.h.d.}$$

the x_i (i = 1,...,n) will be the quasi-cartesian coordinates of the point x_i (i = 1,...,N) for the neighborhood of the point 0,0,0,...,0.

When we were discussing curved space in Sections 18, 19, 20 we were speaking of vectors and tensors; these vectors were vectors of the tangent plane or tangent flat with initial point at the point of contact. We shall not consider any other vectors in connection with curved spaces and we shall refer to these vectors as the vectors of the curved space. To make this idea seem more natural we may remark that a tangent vector to a curve on a surface (or on a curved space) is such a vector, that is, a vector of the tangent plane (or flat) with initial point at the point of contact. In handling these vectors we have been using for the vectors at every point of the curved space a cartesian coordinate system in the flat tangent at that point, in fact, we may say the same system that furnishes us the quasi-cartesian coordinates for the points of the curved space in the neighborhood of the point of contact. We shall, therefore, refer to the cartesian components of the vectors and tensors of the tangent flat when they are considered as vectors and tensors of the curved space - as quasi-cartesian components of these vectors and tensors.

We have thus in connection with every point P on the curved space a local coordinate system which gives quasi-cartesian coordinates of the neighboring points and the quasi-cartesian components of the tensors at P, and in some cases these local coordinate systems are very useful, but it will be necessary to introduce more general, more universal systems and learn how to represent vectors in them. The necessity of this last requirement will be clear if we consider that, although a quasi-cartesian system of a point P may be used to represent points in the neighborhood of P it is not quasi-cartesian for these points and cannot be used as such to represent vectors at such points.

There is no difficulty in introducing a universal system of coordinates for the points of a space - what we want is just as in ordinary space a one-to-one correspondence between the points and n-ples of numbers. A simple example is furnished by the so-called geographical coordinates for the surface of a sphere. Another

approach is given by the so-called parametric representation of a curved space. If the coordinates of a flat space $x_1,\dots,x_N$ are given as functions of one parameter we have a curve; when they are given as functions

25.2 $$x_i(u_1,\dots,u_n) \qquad (i = 1,\dots,N)$$

of n parameters u_j we have what we have called an n-dimensional curved space because eliminating these n parameters from the N equations which express the coordinates in terms of them we find that the coordinates must satisfy N - n = r equations, and this was our definition of an n-dimensional curved space. Now since to every set of values $u_1,\dots,u_n$ of the parameters there corresponds one point of the curved space the parameters u_i may be used as coordinates for the curved space.

Suppose then that we have introduced in some way a general system of coordinates for the points on a curved space (the reader may always think of the special case of a surface). What will be a natural system of representation of vectors to go with it? Just as we use a cartesian system in the tangent flat to represent points on the surface we may, so to say, project the general coordinate system on each tangent flat and use it to represent vectors and tensors in that flat, in particular those with initial points at the point of contact, i.e., the vectors and tensors of the curved space. For the neighborhood of each point we have thus two coordinate systems: the general and the quasi-cartesian for that point - and the same two systems, or, rather their projections, we may use on the corresponding tangent flat. For the neighborhood of each point there will be transformation formulas for the coordinates of points, and from these we can derive transformation formulas for components of vectors and tensors involving the a's, the b's and the g's as introduced in Sections 21, 22 and 23. But since we consider only vectors with initial points at the point of contact we shall use the a's, the b's and the g's calculated from the correspondence between the quasi-cartesian and the general coordinates at a point only for that point itself. We know that the a's and b's are necessary only in the building up of a system so that all we shall need in order to be actually able to handle tensors and vectors in a given general coordinate system are the g's.

We want to explain now how to obtain the g's when a space is given in parametric form 25.2.

We consider a curve $u_i(t)$ on the space, and its tangent vector at some point; it may be considered either as a vector of the curved space, and then its contravariant components will be given, if we denote differentiation with respect to the parameter by a dot placed over a letter, by $\dot u^i$, the square of its length will be $g_{\alpha\beta}\dot u^\alpha \dot u^\beta$, or it may be considered as a vector of the containing space. Its (cartesian) components will be given by $\dot x_i$ $(i = 1,\dots,N)$ where the x_i are functions of t which are obtained from 25.2 by substituting for the u's the expressions $u_i(t)$, characterizing our curve; we have thus

$$\dot x_i = \partial x_i/\partial u_\alpha . \dot u^\alpha$$

and for the square of the vector

$$\sum_{i=1}^{N} \left(\frac{\partial x_i}{\partial u_\alpha}\dot u^\alpha\right)^2 = \sum_{i=1}^{N} \left(\frac{\partial x^i}{\partial u^\alpha}\dot u^\alpha\right)\left(\frac{\partial x^i}{\partial x_\beta}u^\beta\right)$$

$$= \sum_{i=1}^{N} \left(\frac{\partial x_i}{\partial u^\alpha}\cdot\frac{\partial x_i}{\partial u^\beta}\right)\dot u^\alpha \dot u^\beta .$$

Equating this to the expression we obtained above we find

25.3 $$g_{mn} = \sum_{i=1}^{N} \frac{\partial x_i}{\partial u^m}\cdot\frac{\partial x_i}{\partial u^n}.$$

This formula ought to be compared to 21.16 of which it will be seen to be a generalization if account is taken of the values 21.11 of the b's. The method of giving the curved space by means of the formulas 25.11 may be considered as a special case of the one used above; this will be clear if we write 25.11 in the form

$$x_1 = u_1, \quad x_2 = u_2, \quad x_n = u_n,$$

$$x_{n+k} = \tfrac{1}{2}a^{(k)}_{\alpha\beta}u_\alpha u_\beta + \text{t.h.d.}$$

It is seen that the parts of the u's are played by the first n of the x's. Differentiation of the formulas just written with respect to these variables gives

$$\frac{\partial x^i}{\partial u_j} = \delta_{ij} \quad (\substack{i = n\\ j = n}), \qquad \frac{\partial x_{n+k}}{\partial u_j} = a^{(k)}_{\alpha j}x_\alpha + \text{t.h.d.};$$

substitution of these expressions into 25.3 gives

$$g_{mn} = \sum_{i=1}^{n} \frac{\partial x_i}{\partial u_m}\cdot\frac{\partial x_i}{\partial u_n} + \sum_{i=n+1}^{N} \frac{\partial x_i}{\partial u_m}\cdot\frac{\partial x_i}{\partial u_n}$$

25.4 $$= \sum_{i=1}^{n} \delta_{im}\delta_{in} + \sum_{k=1}^{r} a^{(k)}_{\alpha m}a^{(k)}_{\beta n}x_\alpha x_\beta + \text{t.h.d.}$$

$$= \delta_{mn} + \sum_{k=1}^{r} a_{\alpha m}a_{\beta n}x_\alpha x_\beta + \text{t.h.d.}$$

These formulas give the values for the g's in pseudocartesian coordinates for a neighborhood of the point of contact.

For the point of contact itself, i.e., for the origin of our system of coordinates we have

25.41 $$(g_{mn})_o = \delta_{mn}$$

and from the formulas 22.61 we conclude easily that the g's with the upper indices are also the δ's:

25.42 $$(g^{mn})_o = \delta_{mn}$$

As a consequence of this the distinction between covariant and contravariant components vanishes for quasi-cartesian coordinates at the point of contact.

We come now to the operation of differentiation. In the case of flat space we were simply trying to find a system of notations for some operations that were defined independently. Here the situation is different; we have not defined differentiation; we cannot define it in what would seem to be the natural way, as the rate of change of a vector, for instance, because this would necessitate the consideration of the difference between two vectors at two different points and this conception is not defined for curved space.

Before we come to this definition let us formulate the situation in flat space as follows: a tensor field dF has been obtained by differentiation from a tensor field F if in every point the components of dF in a cartesian system are the derivatives of the components of F in that system.

In curved space there is no universal cartesian system but there is a quasi-cartesian system for every point; it is natural, therefore, to define differentiation in curved space by substituting in the above statement "quasi-cartesian system" for "cartesian system"; if we do that we arrive at the following definition:

Definition of Differentiation. We shall say that a tensor field dF has been obtained by differentiation from the tensor field F if at every point the components of dF, in a system of coordinates that is quasi-cartesian at that point, are the derivatives of the components of F in that system of coordinates.

Although this definition may sound complicated it is the simplest imaginable adaptation of the idea of differentiation to curved spaces. The complication arises from the fact that there is no cartesian system in curved space but when we apply this definition to flat space we see that it brings us back to differentiation as we knew it in flat space.

We shall not have actually to pass from general coordinates to quasi-cartesian coordinates and then, after differentiation translate the result back into the language of general coordinates in every special case. We can derive the formulas in general coordinates once for all, just as we did it in the case of flat space in Sections 22 and 23, and we shall obtain exactly the same formulas. The only difference may be in the derivation of the Γ's from the g's (end of Section 22) which was based there on the fact that the derivatives of the g's in cartesian coordinates vanish (22.84). Is this true also in curved space? or, more precisely, do the derivatives of the g's in quasi-cartesian coordinates vanish at the point of contact?

In these coordinates the g's are given by 25.4; differentiating these expressions we obtain

$$25.5 \qquad \partial g_{mn}/\partial x_l = 2\Sigma a^{(k)}_{m\alpha} a^{(k)}_{nl} x_\alpha + \text{t.h.d.}$$

and for the point of contact, where the x's vanish, this is zero so that

$$25.6 \qquad (\partial g_{mn}/\partial x_l)_o = 0.$$

We see thus that formally everything is just the same as in flat space so that we can take over into curved space the whole apparatus of formulas worked out in Sections 21, 22, 23.

Incidentally we may mention that as it follows easily from 25.6 the quantities Γ also vanish in quasi-cartesian coordinates at the point of contact. For future reference we put down the formula

$$25.7 \qquad (\Gamma^i_{jk})_o = 0.$$

In general, the point of contact in quasi-cartesian coordinates is a place where we have the closest possible approach to the situation which obtains in flat space when we use cartesian coordinates.

Another formula that can be easily obtained from 25.5 and that we need later is obtained by differentiating 25.5 once more and setting $x_l = 0$. We get thus

$$25.8 \qquad (\partial^2 g_{mn}/\partial x_i \partial x_j)_o = 2\Sigma a^{(k)}_{mj} a^{(k)}_{ni}.$$

Given a curved space by the formulas 25.2 we know how to find the g's. The question now arises: suppose we are given $\frac{1}{2}n(n+1)$ functions of the coordinates; is it possible to find a space for which these functions serve as the g's. The question is that of solving the system of partial differential equations (25.3), and without going into details we shall state that such a system of equations in general can be solved if the number of unknown functions is equal to that of equations; since we have here $\frac{1}{2}n(n+1)$ equations we must have that many unknown functions; that means that the number of dimensions N of the containing space must be in general $\frac{1}{2}n(n+1)$; in special cases it may, of course, be less than that. We may say then: a two dimensional curved space given by its g's may be always considered as immersed into a 3-dimensional space; a three-dimensional curved space may be always considered as part of a six-dimensional flat space; and a four-dimensional as part of a ten-dimensional flat.

Another question is, whether for given real g's the containing space will come out real; and this is by no means always the case. We know that for $g_{11} = g_{22} = g_{33} = -g_{44} = 1$, all others zero, the minimum cartesian containing space is four-dimensional with one imaginary coordinate and it is clear that no real cartesian space can contain it.

Henceforth we may consider the curved space as given by its g's, and the g's may be considered as arbitrarily given functions of the u's.

It may seem that we lost from view the original purpose of introducing curved space, which was that of obtaining a tensor which we could identify with T_{ij}. We introduced the Riemann tensor having this in mind, but now we seem to be immersed in an entirely formal theory and far removed from the Riemann tensor; as a matter of fact, it is just around the corner; differentiation, although performed according to formulas that are formally the same as in flat space, has, as we shall see, a new content; in trying to discover the difference we will be led to the Riemann tensor from a new point of view.

26. New Derivation of the Riemann Tensor.

We said that the meaning of differentiation in curved space is different from that in flat space. To show this difference in one of its most important manifestations we start out with a tensor of the first rank given in its contravariant components F^i; we differentiate it twice to obtain a tensor of third rank $F^i{}_{,jn}$; in flat space this tensor would not differ from $F^i{}_{,nj}$ because in cartesian components differentiation of a tensor reduces to ordinary differentiation of its components, so that the cartesian components of the two tensors mentioned are $\frac{\partial}{\partial x_n}\frac{\partial f_i}{\partial x_j}$ and $\frac{\partial}{\partial x_j}\frac{\partial f_i}{\partial x_n}$ respectively, and these are equal because the result of ordinary differentiation does not depend on the order; two tensors having equal components in one system of coordinates would be equal in all systems of coordinates and so we have

$$26.1 \qquad F^i{}_{,jn} - F^i{}_{,nj} = 0 \quad \text{in flat space.}$$

This reasoning does not apply to curved space; in fact, to find $F^i{}_{,jn}$ in a point P we have to differentiate $F^i{}_{,j}$; and in order to do that we have to know $F^i{}_{,j}$ in different points of the neighborhood of P; the finding of $F^i{}_{,j}$ in these points involves the use of quasi-cartesian systems for each of these points; we do not have then one quasi-cartesian system in which we can perform all our operations and the reasoning that led us to 26.1 breaks down. In spite of this the result might still hold. In order to show that it does not let us calculate the left hand side of 26.1 using the formulas which we deduced for flat space in Sections 22 and 23 and which, as we proved in Section 25, apply to curved space also.

We start with the contravariant components F_i; we calculate the first differential according to formula 23.6 to obtain

$$26.2 \qquad F^i{}_{,j} = \partial F^i/\partial u_j + \Gamma^i_{\alpha j} F^\alpha;$$

next, we differentiate this again, and get, according to formula 23.7

$$26.21 \qquad F^i{}_{,jn} = \partial F^i{}_{,j}/\partial u_n + \Gamma^i_{\alpha n} F^\alpha{}_{,j} - \Gamma^\beta_{jn} F^i{}_{,\beta};$$

now we form the difference we want to investigate, viz.,

$$F^i{}_{,jn} - F^i{}_{,nj} = \partial F^i{}_{,j}/\partial u_n - \partial F^i{}_{,n}/\partial u_j + \Gamma^i_{\beta n} F^\beta{}_{,j} - \Gamma^i_{\beta j} F^\beta{}_{,n} - (\Gamma^\beta_{jn} - \Gamma^\beta_{nj}) F^i{}_{,\beta};$$

the last bracket vanishes according to 22.71, and what remains becomes after the substitution of the above expression 26.2 for the first differential and rearrangement of terms

$$\frac{\partial^2 F^i}{\partial u_n \partial u_j} - \frac{\partial^2 F^i}{\partial u_j \partial u_n} + \Gamma^i_{\alpha j}\frac{\partial F^i}{\partial u_n} - \Gamma^i_{\beta j}\frac{\partial F^i}{\partial u_n} + \Gamma^i_{\beta n}\frac{\partial F}{\partial u_j} - \Gamma^i_{\alpha n}\frac{\partial F^\alpha}{\partial u_j} + \left(\frac{\partial \Gamma^i_j}{\partial u_n} - \frac{\partial \Gamma^i_n}{\partial u_j} + \Gamma^i_{\beta n}\Gamma^\beta_{\alpha j} - \Gamma^i_{\beta j}\Gamma^\beta_{\alpha n}\right) F^\alpha.$$

Here cancellation takes place in the first three pairs of terms: in the first as a result of independence of ordinary differentiation on order, in the next two pairs as a result of the fact that the name of the index of summation is immaterial; we come out with

$$26.3 \qquad F^i{}_{,jn} - F^i{}_{,nj} = \left(\partial \Gamma^i_{\alpha j}/\partial u_n - \partial \Gamma^i_{\alpha n}/\partial u_j + \Gamma^i_{\beta n}\Gamma^\beta_{\alpha j} - \Gamma^i_{\beta j}\Gamma^\beta_{\alpha n}\right) F^\alpha$$

or

$$26.4 \qquad F^i{}_{,jn} - F^i{}_{,nj} = B^i_{\beta,jn} F^\beta$$

where

$$26.5 \qquad B^i{}_{m,jn} = \partial \Gamma^i_{mj}/\partial u_n - \partial \Gamma^i_{mn}/\partial u_j + \Gamma^i_{\beta n}\Gamma^\beta_{mj} - \Gamma^i_{\beta j}\Gamma^\beta_{mn}.$$

Before we discuss the question whether the expression vanishes we want to show that the B's are the components of a tensor. In fact, multiplying both sides of 26.4 by $X_i Y^j Z^m$, where X_i, Y^j, Z^m are components of arbitrary vectors, and contracting we have

$$(F^\alpha{}_{,\gamma\delta} - F^\alpha{}_{,\delta\gamma}) X_\alpha Y^\gamma Z^\delta = B^\alpha{}_{\beta,\gamma\delta} X_\alpha F^\beta Y^\gamma Z^\delta.$$

The left hand side is a scalar that has been obtained by legitimate operations and is, therefore, independent from the coordinate system used; so is therefore the right hand side, and this proves that the B's are the components of a tensor. (In order to see that this is an essential point and that not every symbol with indices may be considered as a tensor, the reader might consider the expression $\Gamma^\alpha_{\beta\gamma} X_\alpha Y^\beta Z^\gamma$; this expression is, obviously, not independent from the choice of coordinates since in cartesian coordinates the Γ's vanish, and in other systems they do not; the Γ's furnish thus an

example of symbols with indices that can not be interpreted as components of a tensor).

Now we can settle our question as to the vanishing of 26.3 by showing that the B's are mixed components of the Riemann tensor which has been introduced in Section 20. Since we have proved that they are components of a tensor we may use any system of coordinates, and we decide to use a quasi-cartesian system. In such a system the Γ's vanish at the point of contact (25.7) so that we are left with the terms

$$\partial \Gamma^i_{mj}/\partial x_n - \partial \Gamma^i_{mn}/\partial x_j;$$

substituting for the Γ's with one upper index their expressions in terms of the g's with upper indices and the Γ's with all indices down (22.92) we get

$$\partial g^{i\beta}/\partial x_n . \Gamma_{\beta,mj} - \partial g^{i\beta}/\partial x_j . \Gamma^{\beta}_{mn}$$
$$+ g^{i\beta} . \partial \Gamma_{\beta,mj}/\partial x_n + g^{i\beta} . \partial \Gamma_{\beta,mn}/\partial x_j;$$

the first two terms vanish again because the Γ's vanish at the point of contact; taking into account (25.41) that the g's are for the point of contact equal to the δ's, and using the expressions 22.9 we find after a few cancellations

$$B_{im;jn} = \tfrac{1}{2}(\partial^2 g_{ij}/\partial x_n \partial x_m - \partial^2 g_{mj}/\partial x_n \partial x_i$$
$$- \partial^2 g_{in}/\partial x_j \partial x_m + \partial^2 g_{mn}/\partial x_j \partial x_i)$$

(the index i appears here as a subscript because the distinction between contravariant and covariant quantities vanishes for quasi-cartesian coordinates at the point of contact). Using here for the second derivatives of the g's the expressions 25.8 we find

$$B_{im,jn} = \Sigma(a^{(k)}_{in} a^{(k)}_{mj} - a^{(k)}_{ij} a^{(k)}_{mn}).$$

Comparing this to the expression for the Riemann tensor deduced at the end of Section 20 we convince ourselves of the identity of the two expressions.

This shows that, if the Riemann tensor does not vanish, the second differential of a vector field actually may depend on the order of differentiation. This fact is very interesting in itself, it confirms our statement that in curved space differentiation has a new meaning and it has many important implications, on which, however, we cannot dwell here. For us it is important that we have obtained an expression of the Riemann tensor in terms of the g's alone; this means that those properties of the curvature of space which are expressed in the Riemann tensor are determined by the metric of the space, i.e., if distances along different curves are given, the curvature (as far as it is expressed in the Riemann tensor) is determined. According to our conception, the inhabitants of the space certainly can measure lengths; it follows that curvature, as expressed by the Riemann tensor is accessible to the inhabitants, it is an internal property of the space. In particular, for N = 3, n = 2, i.e., for the ordinary surface we obtain the fact that the total curvature can be calculated from the expression for the line element; this is Gauss's Theorema Egregium.

27. Differential Relations for the Riemann Tensor.

The method of quasi-cartesian coordinates in proving a relation between tensors that we used in identifying the B's with the components of the Riemann tensor can be applied often and helps to avoid lengthy computations. We shall use it now to prove certain differential relations for the Riemann tensor that are very important for us because we know that the tensor T^i_j which we want to identify with the contracted Riemann tensor satisfies a certain differential equation, namely, $\partial T^{\alpha}_{\beta}/\partial x_{\alpha} = 0$, and, of course, we expect the tensor in our mathematical theory with which we are going to identify T to satisfy the same relations. In order to deduce differential relations on the contracted Riemann tensor we have to prove first some relations for the non-contracted tensor. These relations have been discovered by Ricci and then rediscovered by Bianchi and bear the latter's name; they are

$$27.1 \qquad R^i_{j;mn,p} + R^i_{j;np,m} + R^i_{j;pm,n} = 0.$$

The proof is very simple if we use quasi-cartesian coordinates. In these coordinates the Γ's at the point of contact vanish and although the first derivatives of the Γ's do not, the components of the tensor obtained by differentiating the B's (formula 26.5) which we have identified with the R's will contain the second derivatives only, because the first derivatives will be multiplied by the Γ's themselves that do vanish. With this remark in mind the proof of the Bianchi relations does not present any difficulty; we simply substitute for each of the three terms in 27.1 the difference of the two second order derivatives and find that the result vanishes identically.

Now, in order to deduce from 27.1 the relations for the contracted tensor we raise in 27.1 the second index so that we have

$$R^{ij}_{\ \ mn,p} + R^{ij}_{\ \ np,m} + R^{ij}_{\ \ pm,n} = 0$$

and here we contract i with m, and j with n. We obtain

$$R^{\alpha\beta}_{\ \ \alpha\beta,p} + R^{\alpha\beta}_{\ \ \beta p,\alpha} + R^{\alpha\beta}_{\ \ p\alpha,\beta} = 0.$$

The second term here may be written as $-R^{\beta\alpha}_{\ \ \beta p,\alpha}$ using the fact (20.71) that the Riemann tensor

changes its sign when two indices of the same pair are interchanged; and the third term is equal to the second as we can see by interchanging α and β (which does not change the value of the expression since α and β are summation indices), and then interchanging the first two indices, i.e., β and α and the next two, i.e., p and β (each of these interchanges changes the sign, so that nothing is changed in the result). We have thus

$$R^{\alpha\beta}{}_{\alpha\beta,p} - 2R^{\alpha\beta}{}_{p\beta,\alpha} = 0.$$

But $R^{i\beta}{}_{j\beta}$ are the mixed components of the contracted Riemann tensor which we denote by $R^i{}_j$ so that, dividing by 2 and changing the sign we have

$$R^{\alpha}{}_{p,\alpha} - \tfrac{1}{2}R^{\alpha}{}_{\alpha,p} = 0.$$

Finally, $R^{\alpha}{}_{\alpha}$ is the result of contraction of the contracted Riemann tensor; we denote this scalar by R (it is called the twice contracted Riemann tensor); then we can write for $R^{\alpha}{}_{\alpha,p}$ simply $R_{,p}$, or $(\delta_{\alpha p}R)_{,\alpha}$ and our relation becomes

$$(R^{\alpha}{}_{p} - \tfrac{1}{2}\delta_{\alpha p}R)_{,\alpha} = 0.$$

28. Geodesics.

In concluding this fragmentary development of the mathematical theory that we intend to apply to Physics in the next chapter we shall study briefly a class of curves in curved space which play an important part in the study of motion. These curves may be considered as generalizations of straight lines in flat space, and we shall begin by considering these.

In agreement with the point of view of differential geometry (Section 21) we shall characterize a straight line by differential equations. If it is given in parametric form (7.11) we obtain by differentiating twice with respect to the parameter and indicating differentiation by a dot placed over the letter

$$\ddot{x}^i = 0. \tag{28.1}$$

Since the choice of the parameter is in a high degree arbitrary, and for another choice of a parameter the representation may cease to be linear and the equations (28.1) may not hold any more - we cannot say that they characterize a straight line; a complete statement would be: a straight line is a curve for which there exists a parametric representation such that 28.1 holds.

Next, we translate 28.1 into the language of curvilinear coordinates, still keeping to flat space. We have, as in Section 21, except that we write now in agreement with Section 23, the index as a superscript,

$$\dot{x}^i = dx^i/dp = \partial x^i/\partial u^\alpha . du^\alpha/dp = b_{i\alpha}\dot{u}^\alpha,$$

$$\ddot{x}^i = d\dot{x}^i/dp = b_{i\alpha}\ddot{u}^\alpha + \partial b_{i\alpha}/\partial u^\beta . \dot{u}^\beta\dot{u}^\alpha = 0.$$

Multiplying by a_{ji} and summing with respect to i, writing $i = \gamma$, we get

$$a_{j\gamma}b_{\gamma\alpha}\ddot{u}^\alpha + a_{j\gamma}.\partial b_{i\alpha}/\partial u^\beta . \dot{u}^\beta\dot{u}^\alpha = 0$$

or, taking into account 21.14 and 22.82

$$\ddot{u}^j + \Gamma^j_{\alpha\beta}\dot{u}^\alpha\dot{u}^\beta = 0. \tag{28.2}$$

We pass now to curved space; in general, we have here no straight lines but we may consider the same equation and investigate the properties of the curves represented by them. We introduce as our definition:

Geodesics are curves which satisfy for an appropriate choice of parameter Equation 28.2.

In studying geodesics it is often more convenient to consider not a single geodesic but a portion of space filled with geodesics, so that through every point there passes one and only one geodesic of the bunch. If we have this situation we have a vector $\ddot{u}^i$ in every point of that portion of space, so that we have a vector field, and the components $\dot{u}^i$ may be considered as functions of the coordinates. We may then write

$$\ddot{u}^j = d\dot{u}^j/dp = \partial\dot{u}^j/\partial u^\beta . du^\beta/dp = \partial\dot{u}^j/\partial u^\beta . \dot{u}^\beta,$$

and equation 28.2 becomes

$$(\partial\dot{u}^j/\partial u^\beta + \Gamma^j_{\alpha\beta}u^\alpha).u^\beta = 0$$

or, according to 23.6,

$$u^j{}_{,\beta}u^\beta = 0. \tag{28.3}$$

This form is very convenient in some cases. We shall use it to prove two properties of geodesics. In the first place we may discuss the meaning of the parameter that we are using. Consider the square of the tangent vector $u^\alpha u_\alpha$; we can prove that this quantity is constant along a geodesic. In fact, differentiating with respect to the parameter, we have

$$\frac{d}{dp}\dot{u}^\alpha\dot{u}_\alpha = \dot{u}^\alpha{}_{,\beta}.\dot{u}^\beta.\dot{u}_\alpha + \dot{u}^\alpha.\dot{u}_{\alpha,\beta}\dot{u}^\beta = 2\dot{u}^\alpha{}_{,\beta}\dot{u}^\beta\dot{u}_\alpha,$$

and this vanishes according to 28.3. Since arc length is given by the formula

$$s = \int\sqrt{\dot{u}^\alpha\dot{u}_\alpha}.dp$$

we see that, as a result of the fact that $\dot{u}^\alpha\dot{u}_\alpha$ is constant, s is proportional to p, or p is proportional to the arc length s. Since multiplication of the parameter by a constant factor will not affect equation 28.2 or 28.3 we may always consider that the parameter is arc length. This discussion does not apply, however, when the quantity $\dot{u}^\alpha\dot{u}_\alpha$ is zero, i.e., when $\dot{u}^i$ is a zero square vector. If we agree to call

curves whose tangent vectors have zero square singular curves we may state the following:

Theorem. In case of a non-singular geodesic, the parameter mentioned in the definition of a geodesic and used in 28.2 and 28.3 is proportional or equal to arc length.

Next we may give an interpretation to equation 28.3 which sheds some light on the geometrical nature of geodesics. We may assume now that in all geodesics of the bunch arc length has been chosen as the parameter; then the vectors $\dot{u}^j$ are unit vectors and they characterize in each point the direction of the geodesic. The derivative $\dot{u}^j{}_{,i}$ characterizes the change of direction as we move in the direction given by the coordinate u^i and $\dot{u}^j{}_{,\alpha}u^\alpha$ gives the change of direction in the direction of the vector $\dot{u}_i$, i.e., in the direction of the geodesic itself. Since, according to 28.3 this quantity is zero we have proved that the direction of a geodesic does not change as we move along it. (The above discussion applies, strictly speaking, only to non-singular geodesics.)

Chapter V.

GENERAL RELATIVITY

In Chapter I we introduced certain fundamental quantities, and we combined them into the symmetric tensor of rank two, T_{ij}. We found that this tensor satisfies the differential equation

$$\partial T_{i\alpha}/\partial x_\alpha = 0,$$

first for $i = 1,2,3$ and then, in Chapter III we showed that, as a result of the new identification introduced there, the fourth equation is also satisfied. We thought it desirable to build a mathematical theory in which a tensor of the same formal properties will appear in a natural way, and in the preceding Chapter IV we succeeded in actually setting up such a theory - the theory of curved space-time.

The structure of such a space, we found, is expressed in a tensor of rank four - the Riemann tensor, but we obtained from it by contraction a tensor of rank two - the contracted Riemann tensor. In investigating the differential properties of the Riemann tensor we found in Section 27 a relation of the type desired; it is satisfied by a tensor which differs slightly from the contracted Riemann tensor, namely, the tensor $R^i_j - \frac{1}{2}\delta_{ij}R$ which we may call the corrected contracted Riemann tensor, and this is the tensor which we are going to identify with the physical tensor T so that our fundamental assumption will be

$$T^i_j = R^i_j - \tfrac{1}{2}\delta_{ij}R.$$

Thus we decide to interpret T, and therefore our fundamental quantities of matter and electricity ρ, u, v, w, X, Y, Z, L, M, N, which went into it, in terms of structure of curved space as it is reflected in the contracted corrected Riemann tensor. But in doing this we find ourselves before a radically new situation. As we wanted, the tensor is now an expression of the properties of space, i.e., the space is now different from the one we had before - geometry and physics is now an organic whole and it is not clear what changes this brings with it; together with the desirable feature, namely the fact that T grew out of space, so to say, we may have brought in some not desirable and hard to manage features. But then there would be no gain if we could merely say that T is a geometrical thing now; we expected to gain something essential in undertaking the merging together of our geometry and physics; and now we stand before an accomplished fact and we have to see what it brought with it. We conjured up something and we do not seem to be able to stop, we have to go ahead and hope that the changes will be beneficial.

It might seem strange that we find a physical interpretation only for the contracted Riemann tensor, only for ten combinations of its twenty components. But this is quite in order. Should all the components of the Riemann tensor be used up in interpreting matter and electricity that would mean that where there is no matter (and electricity) space-time is flat (as far as internal properties are concerned); that would mean that matter acts only where it is; but we know that matter makes itself felt, for instance, by the gravitational field that it produces, also outside the region which it occupies, and this is in accord with our identification as a result of which only part of the components of the Riemann tensor vanish where there is no matter, so that the remaining components may be interpreted as corresponding to gravitational forces.

29. The Law of Geodesics.

In questions of celestial mechanics which we are going to treat now the effects of the electromagnetic field are usually negligible and we shall begin by equating to zero our electromagnetic tensor. Equation 24.9 becomes then

$$T^i_j = -\rho u^i u_j. \tag{29.1}$$

According to our fundamental assumption, this tensor has been identified with the corrected contracted Riemann tensor, and it must satisfy the equation

$$T^\alpha_{i,\alpha} = 0 \tag{29.2}$$

which formally is the same as our old equation of motion 24.8 but differs from it in that it has to be interpreted in curved space. The last two equations impose certain conditions on the velocity components u^i and we want to find these conditions or, in other words, we want to eliminate density from the equations 29.1, 29.2. (We have been using in Chapter IV the letter u for the general coordinates - in this chapter we go back to our notation of the first three chapters and denote by u^i again the four-dimensional velocity vector, and we shall denote the general coordinates by x^i.)

First of all we shall prove the following theorem due to Mineur.

Theorem. If the field u^i satisfies the equations 29.1, 29.2 the vectors u^i may be considered as tangent unit vectors to a family of geodesics filling the space.

Proof. We consider first the case when u_i is a unit vector (and not a zero square vector), i.e., $u_\beta u^\beta = -1$. Differentiating this relation we have

29.3 $$u_{\beta,i}u^\beta = 0.$$

Substituting 29.1 into 29.2 we get

29.4 $$\partial\rho/\partial x_\alpha . u^\alpha u_j + \rho u^\alpha{}_{,\alpha} u_j + \rho u^\alpha u_{j,\alpha} = 0.$$

Dividing by ρ and introducing the notation

29.5 $$A = \partial \log \rho/\partial x_\alpha . u^\alpha + u^\alpha{}_{,\alpha}$$

we may write 29.3 as

$$Au_j + u^\alpha u_{j,\alpha} = 0.$$

Multiplying this by u^j and summing with respect to j, for which we write β, we have

$$Au^\beta u_\beta + u^\alpha u_{\beta,\alpha} u^\beta = 0$$

which, according to 29.3 gives $A = 0$. Substituting this into 29.5 we obtain

29.6 $$u^\alpha . u_{j,\alpha} = 0$$

which, according to 28.3 proves the theorem in the case considered.

But we also have to consider the case of propagation of light. In this case we do not have to consider any density ρ the momentum vector being given by q_i with $q_\beta q^\beta = 0$. The preceding proof breaks down in this case, but continuity considerations lead us to the result that, in this case also we can find a scalar field ρ such that q^i/ρ will be tangent vectors to geodesics.

We conclude that in a gravitational field matter and light particles follow geodesics.

In the present chapter we are going to apply this result to the investigation of the motion of a planet and the propagation of light in the Solar system. We shall see that the changed significance of differentiation takes care, in a way, of what is usually accounted for by gravitational forces.

30. Solar System. Symmetry Conditions.

Our equations 29.1 and 29.2 describe relations existing between matter and field. We proved that the motion of matter is characterized by the geodesics of the curved space, but the curvature is in turn determined by matter. Theoretically, we may have a complete description of the situation, but in practice we do not know how to handle it, we do not know where to begin. In such cases we often resort to the method of successive approximations. Let us try to apply this method here. In investigating the motion of a planet around the sun we neglect in the first place the motion of the sun. Then, in the first approximation we neglect the mass of the planet, i.e., assume that there is no matter outside the sun. Since we have already neglected electromagnetism we have then that outside the sun the tensor T is zero so that, according to the fundamental assumption,

$$R^i_j - \tfrac{1}{2}\delta_{ij}R = 0.$$

Contracting we get $R - \tfrac{1}{2}.4R = 0$, so that $R = 0$ and we have simply

30.1 $$R^i_j = 0.$$

These equations are known as Einstein's equations. We see that the statement that the corrected contracted Riemann tensor vanishes is equivalent to the statement that the contracted Riemann tensor vanishes.

As a result of our first approximation we derived thus the field equations 30.1. In the next approximation we introduce the planet and assume that its action on the field is negligible but that the field acts on it, i.e., that the motion of the planet is given by the geodesics of the field which has been determined in the preceding step; the motions will then be given by the equations 28.2

30.2 $$\ddot{x}^i + \Gamma^i_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta = 0$$

in which the Γ's are calculated from the g's which have been found to satisfy 30.1.

Our problem, therefore, falls in two: first, to find a field satisfying the equations 30.1, and second, to find the geodesics of this field.

In this form the problem is comparable to the problem in Newtonian mechanics as explained in Section 1. There the field was given by the potential which had to satisfy the Laplace equation; here the field is given by the g's which have to satisfy the equations 30.1.

There the motion, after the field had been determined, was described by second order ordinary equations, differentiation being taken with respect to time; here motion is also described by second order differential equations, derivation being with respect to s.

It is possible by making some special assumptions, neglecting certain quantities, for instance the derivatives of all the g's except g_{44} and dropping some terms, to obtain the general Newtonian equations as a special or limiting case of our equations. The equations 30.1 would thus reduce to the Laplace equation 1.54 for g_{44} and the equations of a geodesic to the equations of motion 1.1 in which X,Y,Z are given by 1.53, so that we could consider the general Newtonian theory of motion in a gravitational field as a first approximation to the theory of Relativity, but it is quite difficult in the general case to estimate what we neglect and the error we commit, and we prefer to compare the two theories on some concrete special

cases. All these cases will refer to what corresponds to a gravitational field produced by a single attracting center. We found in Section 1 such a field by using the general equations and, in addition, the condition of symmetry. We intend to follow an analogous course here. Our general equations are 30.1 and now we want to find what will correspond to the conditions of symmetry. The situation is much more complicated here. There the field could be characterized by a scalar φ and the condition of symmetry with respect to a point was simply expressed by stating that φ is a function of distance from that point; here the field is characterized by a tensor g_{ij}. There, in the second place, we worked in ordinary space; here we have space-time which has an additional coordinate, t. Last, there the space was given and in it distances were well defined; on this space was superimposed a field whose symmetry we had to discuss; here the field is not superimposed on a space with a given metric, but the metric itself constitutes a field which has to be determined by the symmetry condition.

We shall take up these three difficulties one by one.

In the first place let us consider a tensor field in ordinary space, and let us impose on it the condition of symmetry with respect to a point. A tensor we may consider (Section 9) as the left hand side of an equation of a central quadric surface (we are interested in a symmetric tensor here, since the g's are symmetric in the indices i and j - this symmetry we must try not to confuse with the symmetry with respect to a point which we impose on the field - and a symmetric tensor is sufficiently characterized by a quadratic form) which we may consider as an ellipsoid. Our tensor field will then be represented by an ellipsoid at every point of space. The field must allow rotations around a fixed center O, i.e., such a rotation must bring the field into itself; in other words, if a rotation brings a point P into a point Q it must bring the ellipsoid at P into the ellipsoid at Q. In particular, a rotation, which leaves P unchanged must not change the ellipsoid at P. It is clear that every ellipsoid must be an ellipsoid of revolution and that its axis must be directed along the radius vector from O to P.

The ellipsoid at the point x,0,0, will be seen to be

$$A(\xi - x)^2 + B(\eta^2 + \zeta^2) = 1$$

and for a general point P, if we use polar coordinates for P and (considering if it helps, the ellipsoid as infinitesimal) their differentials for the points of the quadratic relative to P,

$$30.3 \qquad Adr^2 + B(d\theta^2 + \sin^2\theta . d\varphi^2) = 1.$$

Since ellipsoids at points equidistant from O must have the same dimensions, A and B must be functions of r alone.

The left hand side of this equation gives a tensor field which satisfies the condition of symmetry with respect to a point. Next, we consider the complication resulting from the introduction of time. In Section 1 time was not mentioned, it means that the field was considered as independent of time, or static; we may say that the field must not be affected by a change in t or, from the four-dimensional point of view, by a translation along the t-axis. This is a requirement of the same character as that of symmetry with respect to a point; from the four-dimensional point of view we may combine the two requirements and say that the field must be symmetric with respect to a line - the t-axis. But the field now is a field in four-space, it will be represented by a quadratic form in dr, $d\theta$, $d\varphi$ and dt. For $dt = 0$ it must reduce to the field given before; the coefficients must be independent of t corresponding to the requirement that the field be static; and a change from t to -t must also not affect the field (reversibility of time) so that terms of the quadratic form involving dt to the first power must be absent. It follows that the addition of the fourth dimension results in the addition of only one term to our tensor which now may be written as

$$30.4 \qquad Adr^2 + B(d\theta^2 + \sin^2\theta . d\varphi^2) + Cdt^2$$

where C, as well as A and B, are functions of r alone.

And now we have to overcome the last difficulty, that connected with the fact that our space is curved and that we cannot define symmetry in terms of rotations because rotation means a transformation in which distances are preserved, and distances are defined only by the field of the g's which we want to determine by the requirement that it be not affected by rotations. To overcome this difficulty we have to agree on some other definition of symmetry, and it seems natural to adopt as such the following: in order to define a symmetry for a curved space we shall compare it with a flat space by establishing a one-to-one correspondence between the points of the two spaces. Corresponding to every transformation of the flat space we will have then a transformation of the curved space; and we shall say that the curved space possesses the same symmetry as a field F in the flat space if the metric of the curved space - as given by the g's - is not affected by those transformations of the curved space which correspond to the transformations in flat space not affecting the field F.

Suppose now that we have such a curved space. This implies that we have a one-to-one correspondence with the flat space, and we may

use for the points of the curved space the same coordinates that we use for the corresponding points of the flat space. It is clear that 30.4 will satisfy the requirements, so that we can take it for our fundamental tensor, or as we shall say (compare 21.8) for our ds^2.

But the quantities r, θ, φ, t, which have definite geometrical significance in flat space lose it in curved space - they are just numbers which we use to characterize different points as we use numbers to characterize houses on a street. There is no reason why we should not replace them by other numbers, i.e., transform our coordinates, if it would simplify our formulas. Now, it is clear that transformations involving θ, φ, t will make our expression 30.4 more complicated because it would introduce these coordinates into the coefficients. But we could choose a transformation on r alone so as to simplify that expression; we could, for instance, reduce any one coefficient to a prescribed function of the new r. We make this choice in such a way as to reduce B to r^2 because, in a way, it restores to r a geometrical meaning as we shall see presently. If we write $\xi(r)$ and $-\eta(r)$ for the functions of the new r which now appear instead of A and C, and interpret 30.4 as giving $-ds^2$, in accordance with the standardization of the parameter adopted in Section 12, our final formula will be

30.5

$$-ds^2 = \xi(r)dr^2 + r^2(d\theta^2 + \sin^2\theta.d\varphi^2) - \eta(r).dt^2.$$

Letting here r and t have constant values we have a surface, and a simple calculation would show that $\frac{1}{r^2}$ is the total curvature of this surface, which gives a geometrical meaning to r.

Our task is now accomplished, we have imposed on our space the conditions of symmetry and we have next to impose on it the general equations 30.1.

31. Solution of the Field Equations.

We are now at a stage which corresponds to the assumption that the potential φ is a function of r alone in Section 1, and our next task corresponds to the substitution of $\varphi(r)$ into Laplace's equation. Instead of one unknown function $\varphi(r)$ we have here the ten g's determined by 30.5 which we may write out as

$$g_{11} = \xi(r), \quad g_{22} = r^2,$$

31.1

$$g_{33} = r^2\sin^2\theta, \quad g_{22?} = -\eta(r)$$

all others zero,

and which involve two unknown functions. In order to determine these functions we have to substitute 31.1 into 30.1. In the first place we have to calculate the g's with the upper indices from the formulas (22.6)

$$g_{i\alpha}g^{\alpha j} = \delta_{ij}.$$

Since the g's with two distinct lower indices vanish, only those terms on the left are not zero in which $\alpha = i$ and we have

$$g_{ii}g^{ij} = \delta_{ij};$$

for $i \neq j$ the right hand sides are zero, and since the first factors on the left are not zero the second must vanish; we see thus that the g's with two distinct upper indices also vanish. For $j = i$ we have unity on the right and thus

$$g^{11} = 1/\xi(r) \qquad g^{22} = 1/r^2,$$

31.2

$$g^{33} = 1/r^2\sin^2\theta, \quad g^{44} = -1/\eta(r)$$

all others zero.

In what follows

$$x_1 = r, \quad x_2 = \theta, \quad x_3 = \varphi, \quad x_4 = t.$$

Differentiation with respect to r will be denoted, as in Section 1, by '. We next calculate the Γ's with all indices down according to 22.9 and obtain, omitting those that come out zero,

$$\Gamma_{1,11} = \tfrac{1}{2}\xi', \quad \Gamma_{1,22} = -r, \quad \Gamma_{1,33} = -r\sin^2\theta,$$

$$\Gamma_{1,44} = -\tfrac{1}{2}\eta', \quad \Gamma_{2,33} = -r^2\sin\theta.\cos\theta,$$

31.3

$$\Gamma_{2,12} = r, \quad \Gamma_{3,13} = r\sin^2\theta,$$

$$\Gamma_{3,23} = r^2\sin\theta.\cos\theta, \quad \Gamma_{4,14} = -\tfrac{1}{2}\eta'.$$

Raising of an index is accomplished in this case simply by multiplying by the g with the index to be raised appearing above twice, because the sum $g^{i\alpha}F_\alpha$ which, according to 23.3 is equal to F^i reduces, as the result of the vanishing of the g's with two equal upper indices to one term, namely $g^{ii}F_i$. This permits us to write out easily the Γ's with one index above:

$$\Gamma^1_{11} = \frac{\xi'}{2\xi}, \quad \Gamma^1_{22} = -\frac{r}{\xi}, \quad \Gamma^1_{33} = -\frac{r\sin^2\theta}{\xi},$$

31.31 $$\Gamma^4_{44} = -\frac{\eta'}{2\xi}, \quad \Gamma^2_{33} = -\sin\theta.\cos\theta,$$

$$\Gamma^2_{12} = \frac{1}{r}, \quad \Gamma^3_{13} = \frac{1}{r}, \quad \Gamma^3_{23} = \cot\theta, \quad \Gamma^4_{14} = \frac{\eta'}{2\eta}.$$

Next we have to calculate those components of the Riemann tensor which appear in the

expressions for the components of the contracted Riemann tensor, i.e., those with the first index equal to the one before last or those of the type $R^{i}{}_{j;ih}$. We do not write those out but state that the result of the calculation with their aid of the components of the contracted Riemann tensor is, that all these components with two distinct indices vanish and the others are

$$R_{11} = \frac{\xi'}{\xi r} + \frac{\xi'\eta'}{4\xi\eta} + \frac{\eta'^2}{4\eta^2} - \frac{\eta''}{2\eta}$$

$$R_{22} = \frac{\xi' r}{2\xi^2} + \left(1 - \frac{1}{\xi}\right) - \frac{r\eta'}{2\xi\eta}$$

$$R_{33} = R_{22} \sin\theta$$

$$R_{44} = -\frac{\xi'\eta'}{4\xi^2} - \frac{\eta'^2}{4\eta\xi} + \frac{\eta''}{2\xi} + \frac{\eta'}{\xi r}.$$

It is more convenient to operate with the mixed components of the contracted Riemann tensor (although it is not necessary, and the reader might for the sake of practice go through the same calculations using covariant components), and these are obtained from the last formulas by multiplication by the corresponding g with upper indices; we obtain thus

$$R^{1}_{1} = \frac{\xi'}{\xi^2 r} + \frac{\xi'\eta'}{4\xi^2\eta} + \frac{\eta'^2}{4\xi\eta^2} - \frac{\eta''}{2\xi\eta}$$

31.4
$$\left.\begin{aligned} R^{2}_{2} &= \\ R^{3}_{3} &= \end{aligned}\right\} \frac{\xi'}{2\xi^2 r} + \frac{1}{r^2}\left(1 - \frac{1}{\xi}\right) - \frac{\eta'}{2\xi\eta r}$$

$$R^{4}_{4} = \frac{\xi'\eta'}{4\xi^2\eta} + \frac{\eta'^2}{4\eta^2\xi} - \frac{\eta''}{2\xi\eta} - \frac{\eta'}{\xi\eta r}.$$

We come now to the ten equations $R^{i}_{j} = 0$ that we have to satisfy; six of them, namely, those in which $i \neq j$ are satisfied identically because our R's as well as the δ's vanish for distinct indices. Of the remaining four equations the second and the third are identically the same because of the equality of the corresponding values of R in 31.4. Three equations remain, viz.,

$$\frac{\xi'}{\xi^2 r} + \frac{\xi'\eta'}{4\xi^2\eta} + \frac{\eta'^2}{4\xi\eta^2} - \frac{\eta''}{2\xi\eta} = 0$$

31.5
$$\frac{\xi'}{2\xi^2 r} + \frac{1}{r^2}\left(1 - \frac{1}{\xi}\right) - \frac{\eta'}{2\xi\eta r} = 0$$

$$\frac{\xi'\eta'}{4\xi^2\eta} + \frac{\eta'^2}{4\eta^2\xi} - \frac{\eta''}{2\xi\eta} - \frac{\eta'}{\xi\eta r} = 0$$

Subtracting the last one from the first we have

31.6
$$\frac{1}{\xi r}\left(\frac{\xi'}{\xi} + \frac{\eta'}{\eta}\right) = 0$$

whence

$$\xi\eta = \text{const.} = c.$$

By choosing our unit of time appropriately we can reduce this constant to 1, so that

31.7 $\qquad \xi\eta = 1 \qquad \text{or} \qquad \xi = \frac{1}{\eta}.$

Using 31.6 and 31.7 in the second of the equations 31.5 we obtain

$$\eta' r = 1 - \eta$$

which gives

31.8 $\qquad \eta = 1 - \gamma r$

where γ denotes a constant of integration.

Our field then is given by

31.9 $\qquad -ds^2 = \frac{dr^2}{\eta} + r^2 d\theta^2 + r^2 \sin^2\theta . d\varphi^2 - \eta dt^2$

where η is given by 31.8.

32. Equations of Geodesics.

We first consider the non-zero geodesics which correspond to a material particle. We know that in this case arc-length can be taken as parameter so that the curve in addition to the equations 30.2 must satisfy the equation 30.5 which we may write as

32.1 $\qquad \frac{\dot{r}^2}{\eta} + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\varphi}^2 - \eta\dot{t}^2 = -1;$

we shall, however, make our discussion slightly more general and write A in the right hand side with a view of using the results also in the case corresponding to a light particle. We shall discuss this equation together with the equations 30.2 which become here

32.21 $\qquad \ddot{r} - \frac{\eta'}{2\eta}.\dot{r}^2 + r\eta.\dot{\theta}^2 + r\eta\sin^2\theta.\dot{\varphi}^2 + \frac{\eta\eta'}{2}.\dot{t}^2 = 0$

32.22 $\qquad \ddot{\theta} + 2\frac{\dot{r}}{r}.\dot{\theta} - \sin\theta\cos\theta\,\dot{\varphi}^2 = 0,$

32.23 $\qquad \ddot{\varphi} + 2\frac{\dot{r}}{r}.\dot{\varphi} + 2\cot\theta\,.\dot{\theta}\dot{\varphi} = 0,$

32.24 $\qquad \ddot{t} + \frac{\eta'}{\eta}.\dot{r}\dot{t} = 0.$

The choice of the θ and φ coordinates is at our disposal. We choose them in such a way that the initial position of the particle be on the equator and that the tangent be tangent to the equator. In this case $\theta = \frac{\pi}{2}$ and $\dot{\theta}$ = zero at the initial moment and the second equation shows that $\theta = \frac{\pi}{2}$ always. Now the last two equations may be integrated once each and they furnish

32.3 $\qquad \dot{\varphi}r^2 = h$

32.4 $\qquad \dot{t}\eta = k$

where h and k are constants. Together with these two equations we have to consider the one corresponding to 32.1; viz.,

$$32.5 \qquad \frac{\dot r^2}{\eta} + r^2\dot\varphi^2 - \eta\dot t^2 = A.$$

We simplify our system of equations in the following way: (a) we eliminate t by means of 32.4; (b) we eliminate differentiation with respect to the parameter by using $\dot r = (dr/d\varphi).\dot\varphi$ and 32.3; (c) we introduce as a new variable, as is customary in celestial mechanics, the inverse distance $u = \frac{1}{r}$, instead of r, so that

$$32.6 \qquad r = \frac{1}{u};$$

and, (d) we substitute the value for η from 31.8. We obtain in this way a differential equation between u and φ; viz.,

$$32.7 \qquad \left(\frac{du}{d\varphi}\right)^2 + u^2 = \lambda - A\frac{\gamma}{h^2}.u + \gamma u^3$$

where λ is a constant. This equation may be considered as the equation of the orbit of a planet.

33. Newtonian Motion of a Planet.

Every reader knows, of course, that according to the Newtonian theory a planet moves around the sun on an ellipse in one of whose foci the sun is situated, although he may not be in the possession of a proof of that; we shall not give a proof of that here either, but we shall discuss in detail only one feature of the situation. The vertex of the ellipse which is nearest to the focus in which the sun is located is called the perihelion, the other vertex - the aphelion; the line joining the perihelion and the aphelion is the major axis, and therefore passes through the sun. Using the coordinates u and φ corresponding to those of the preceding section we may say that the perihelion corresponds to the maximum value of u, and the aphelion to the minimum value of u, and that the transition from the maximum to the minimum value of u corresponds to the change of φ by the amount π. It is this last fact that we shall deduce from the equations of motion. We may (corresponding to the fact that we set $\theta = \frac{1}{2}\pi$ in the preceding section) consider a motion in the xy-plane characterized by the equations (see 1.1 and 1.3)

$$33.1 \qquad \frac{d^2x}{dt^2} + \frac{Mx}{r^3} = 0, \qquad \frac{d^2y}{dt^2} + \frac{My}{r^3} = 0.$$

We have now to introduce variables corresponding to those used above in the Relativity treatment, i.e., to set

$$x = \cos\varphi/u, \qquad y = \sin\varphi/u.$$

We calculate the first and the second derivatives of x and y with respect to t, substitute them into 33.1 and combining the terms with $\cos\varphi$ and those with $\sin\varphi$ we get

$$\cos\varphi\left[2u^{-3}\left(\frac{du}{d\varphi}\right)^2 - u^{-2}\frac{d^2u}{dt^2} - u^{-1}\left(\frac{d\varphi}{dt}\right)^2 + Mu^2\right]$$

$$+ \sin\varphi\left[2u^{-2}\frac{du}{dt}.\frac{d\varphi}{dt} - u^{-1}\frac{d^2\varphi}{dt^2}\right] = 0,$$

$$\sin\varphi\left[2u^{-3}\left(\frac{du}{d\varphi}\right)^2 - u^{-2}\frac{d^2u}{dt^2} - u^{-1}\left(\frac{d\varphi}{dt}\right)^2 + Mu^2\right]$$

$$- \cos\varphi\left[2u^{-2}\frac{du}{dt}.\frac{d\varphi}{dt} - u^{-1}\frac{d^2\varphi}{dt^2}\right] = 0.$$

Multiplying the first of these equalities by $\sin\varphi$, the second by $\cos\varphi$ and adding the results we obtain

$$33.2 \qquad 2u^{-3}\left(\frac{du}{dt}\right)^2 - u^{-2}\left(\frac{d^2u}{dt^2}\right) - u^{-1}\left(\frac{d\varphi}{dt}\right)^2 + Mu^2 = 0$$

and then easily

$$2u^{-2}.\frac{du}{dt}.\frac{d\varphi}{dt} - u^{-1}.\frac{d^2\varphi}{dt^2} = 0.$$

The last equation may be written as

$$\frac{d^2\varphi/dt^2}{d\varphi/dt} = \frac{2du/dt}{u}$$

whence

$$33.3 \qquad d\varphi/dt = Hu^2$$

where H is a constant of integration. We next want to eliminate t from 33.2 with the help of the last formula. Differentiating it we have

$$\frac{d^2\varphi}{dt^2} = 2Hu.\frac{du}{dt} = 2Hu.\frac{du}{d\varphi}.\frac{d\varphi}{dt} = 2H^2u^3.\frac{du}{d\varphi}$$

and then

$$\frac{du}{dt} = \frac{du}{d\varphi}.\frac{d\varphi}{dt} = \frac{du}{d\varphi}.Hu^2$$

$$\frac{d^2u}{dt^2} = \frac{d^2u}{d\varphi^2}\left(\frac{d\varphi}{dt}\right)^2 + \frac{du}{d\varphi}.\frac{d^2\varphi}{dt^2} = \frac{d^2u}{d\varphi^2}H^2u^4 + 2H^2u^3\left(\frac{du}{d\varphi}\right)^2$$

Substituting into 33.2 we arrive at

$$2u^{-3}\left(\frac{du}{d\varphi}\right)^2H^2u^4 - u^{-2}\left[\frac{d^2u}{d\varphi^2}H^2u^4 + 2H^2u^3\left(\frac{du}{d\varphi}\right)^2\right]$$

$$-u^{-1}H^2u^4 + Mu^2 = 0$$

and, after two terms cancel, at

$$33.4 \qquad \frac{d^2u}{d\varphi^2} + u = \frac{M}{H^2}.$$

This, we easily see, may be obtained by differentiation from

$$33.5 \qquad \left(\frac{du}{d\varphi}\right)^2 + u^2 = \alpha + \frac{2Mu}{H^2}$$

where α is a constant. This corresponds to equation 32.7 obtained from Relativity theory in the preceding section; in that last equation we have, of course, to take $A = -1$ if we consider the motion of a planet so that it becomes

$$33.51 \qquad \left(\frac{du}{d\varphi}\right)^2 + u^2 = \lambda + \gamma\frac{u}{h^2} + \gamma u^3,$$

and we see that the difference is essentially in only one term. But before we come to the comparison of the motions described by these two equations we have to continue the discussion of 33.5. The character of motion described by it depends on the values of the constants appearing in it, and also on the initial conditions. We begin the discussion by writing 33.5 in the form

$$33.52 \qquad \left(\frac{du}{d\varphi}\right)^2 = -(u - u_1)(u - u_2)$$

where u_1 and u_2 are the two roots of the polynomial $u^2 - 2Mu/H^2 - \alpha$. If the two roots are complex, or equal, the right hand side of 33.52 is negative and we cannot have real motion. Also when both roots are negative the right hand side is negative for positive values of u (and u, being the inverse distance, must be positive). The case of one positive and one negative root corresponds to u changing from 0 to a finite value and then going back to zero, for instance, a comet approaching the sun from an infinite distance and then receding back into infinity. But we want to treat the case of a planet, and this will obviously correspond to the only remaining case; viz., that of two distinct positive roots. If by u_1 we denote the larger and by u_2 the smaller of the two roots it will be convenient to write our equation as

$$33.53 \qquad \left(\frac{du}{d\varphi}\right)^2 = (u_1 - u)(u - u_2),$$

and we see that a real solution is only possible when u is between u_2 and u_1. The motion will manifest itself in an oscillation of u between u_2 and u_1 and the sign of $du/d\varphi$ will change at these points. The particular question we want to investigate is, as was mentioned at the beginning of the section, to what change of φ corresponds one oscillation of u, between u_2 and u_1, say. In order to find this we solve the equation for $d\varphi$, obtaining

$$d\varphi = \frac{du}{\sqrt{(u_1 - u)(u - u_2)}},$$

whence

$$33.6 \qquad \varphi_1 - \varphi_2 = \int_{u_2}^{u_1} \frac{du}{\sqrt{(u_1 - u)(u - u_2)}};$$

a change of variable will help us to evaluate this integral. We put

$$33.7 \qquad \frac{u - u_2}{u_1 - u_2} = \sin^2 x;$$

when x changes from 0 to $\pi/2$, u will increase from u_2 to u_1 as required. We have

$$33.8 \qquad \begin{aligned} du &= 2(u_1 - u_2)\sin x \cos x\, dx, \\ u - u_2 &= (u_1 - u_2)\sin^2 x \\ u_1 - u &= u_1 - \left[u_2 + (u_1 - u_2)\sin^2 x\right] \\ &= (u_1 - u_2)\cos^2 x, \end{aligned}$$

and the integral becomes

$$33.9 \qquad \varphi_1 - \varphi_2 = \int_0^{\frac{\pi}{2}} 2dx = \pi.$$

The answer to our question is then, that φ changes exactly through π while u performs an oscillation between its minimum and its maximum values, which corresponds to the fact mentioned before that the aphelion and the perihelion are on a straight line with the sun, which fact we thus proved.

34. Relativity Motion of a Planet.

Following this excursion into celestial mechanics according to Newton we return to our Relativity formulas which we shall treat by comparing them to the formulas derived in the last section.

At this stage we come again upon a fundamental question: we have two theories; the quantities of one of them have been identified with measured quantities, and this identification proved, in the main, a splendid success; if the new theory is to be applied successfuly, it is clear that it has essentially to agree with the old theory with which it may be compared instead of being compared with results of measurement directly; that means that we have to establish a correspondence between quantities of the two theories, and we have to expect that the corresponding quantities of the two theories obey approximately the same relations. This correspondence has been anticipated in the preceding pages by using the same letters for quantities which it is intended to identify. But it may not be superfluous to remind the reader that the quantities u, φ, θ of the two theories are not the same; there is a certain arbitrariness in choosing coordinates in curved space, and especially obvious it must be in the case of r (of which u is the inverse); it is possible to substitute for r some simple function of r, and, indeed, it has been done; the criterion of correctness of choice must lie in the success of the identification.

Next we must identify the constants of the new theory with those of the old. It would seem as though we must, in order to reach an agreement, make $\gamma = 0$ so as to get rid of the last term of the equation 33.51 by which it differs essentially from 33.5. But this would annihilate also the preceding term in the new formula and so spoil the correspondence altogether. We must, therefore, ascribe to γ a finite value, but we will expect that it will be small; more precisely, it will be small in such a way that the term γu^3 will not affect the equation 33.51 essentially or will be small in comparison with u^2. Next, let us compare 32.3 with 33.3. Of course, the left hand sides differ by the factor dt/ds, but this is equal (Section 13) to $1/\sqrt{1-\beta^2}$ which is, even for motions of planets, very close to one, so that, in the first approximation we may identify h with H. Comparing now 33.5 and 33.51 we come to the conclusion that

34.1 $$\gamma = 2M$$

so that we may write 35.51 as

34.2 $$\left(\frac{du}{d\varphi}\right)^2 + u^2 = \alpha + \frac{2Mu}{H^2} + 2Mu^3.$$

After we have made these identifications the situation is then this: if we neglect the term $2Mu^3$ in the equation, and this term is negligible in most cases, we have the same equation of the orbit as in Newtonian mechanics. This result is very satisfactory, we have been able to obtain the equations of motion of a planet without considering any gravitational forces, as a result of our identification of the contracted Riemann tensor with the complete tensor. Still the term $2Mu^3$ is there, the Relativity theory predicts an orbit that is slightly different from that predicted by the Newtonian theory; is the difference within the error of observation? We shall consider now this question, but instead of considering the motion as a whole, we shall consider only the feature of it which for the case of Newtonian motion has been considered in the preceding section; viz., we shall ask ourselves whether, corresponding to an oscillation of u between a minimum and a maximum value, the change in φ will be exactly π. Of course, we are sure that in the new theory there will be motions which differ but slightly from the motion considered in the preceding section, so that the general character of the motion will be the same, and u will oscillate between a minimum and maximum. The value of $du/d\varphi$ will now be expressed by a polynomial of degree <u>three</u> the first two terms of which are

$$2Mu^3 - u^2.$$

The sum of the three roots of this polynomial is $1/2M$ so that if u_1, u_2 denote two roots, viz., those two roots which differ but slightly from the roots denoted in the same way in the preceding case, the third root will be

$$\frac{1}{2M} - u_1 - u_2,$$

and the integral corresponding to 33.6 will be

$$\int_{u_2}^{u_1} \frac{du}{\sqrt{(u_1 - u)(u - u_2)\left[1 - (u_1 + u_2 + u)2M\right]}};$$

the same substitution 33.7 as before will be applied. We only have to calculate

$$u_1 + u_2 + u - u_1 + u_2 + u_2 + (u_1 - u_2)\sin^2 x$$

$$= u_1 + u_2 + u_1 \sin^2 x + u_2 \cos^2 x,$$

so that the integral becomes

$$\varphi_1 - \varphi_2 = \int_0^{\frac{\pi}{2}} \frac{2dx}{\sqrt{1 - 2M(u_1 + u_2 + u_1 \sin^2 x + u_2 \cos^2 x)}}.$$

As we saw before, M is a very small quantity; before, we neglected it altogether and obtained π for the value of the integral; now, we shall go to the next approximation; we shall develop the denominator according to the powers of M and neglect all terms beyond the second (it would be a very easy but not a worthwhile matter to estimate the value of the error); we get in this way, as an approximate value for $\varphi_1 - \varphi_2$

$$\int_0^{\frac{\pi}{2}} 2\left[1 + M(u_1 + u_2 + u_1 \sin^2 x + u_2 \cos^2 x)\right] dx,$$

or

$$\pi + \frac{3\pi M}{2}(u_1 + u_2).$$

The new theory predicts then that the angle φ will have changed by this amount while the distance from the sun changes from its minimum to its maximum; i.e., that the perihelion and aphelion are not in a straight line with the sun but that the planet moves through an additional angle of $\frac{3\pi}{2}(u_1 + u_2)$ after reaching the position opposite the one where it was during the perhelion, before reaching the aphelion. Since the same situation applies to the motion between an aphelion and the next perihelion we see that between two consecutive perihelia the planet will have moved through an angle $2\pi + 3\pi M(u_1 + u_2)$, or that the perihelion will have moved through an angle $3\pi M(u_1 + u_2)$ during one revolution of the planet. This is a very small amount, and it may be considered as a correction to the classical result according to which the planet moves on an elliptic orbit with the sun in one of the foci. If a is the major semi-axis and e the eccentricity, the distance at perihelion is a - ae and the distance at aphelion a + ae: we have then

$$u_1 + u_2 = 1/(a - ae) + 1/(a + ae) = 2/a(1 - e^2),$$

and the final formula for the advance of the perihelion comes out

$$34.5 \qquad p = \frac{6\pi M}{a(1 - e^2)}.$$

Here then we have two predictions: on the old theory the perihelion will remain fixed in space; according to the new one it will advance by p during one revolution. What are the results? In the case of most planets either this amount is too small or the position of the perihelion too uncertain to permit any decision but in the case of the planet Mercury it was known for a long time that there is a discrepancy between the prediction of the Newtonian theory and actual observations; and it happens that the discrepancy is very nearly the amount showing the discrepancy between the two theories, so that the theory of Relativity predicts a result that has been actually observed. This must be considered as a success of the new theory.

35. Deflection of Light.

According to Section 29 a light particle also moves along a geodesic, only in this case it is a zero geodesic, one along which the tangent vectors have zero length. The equations for such a geodesic are the same as for the other kind with the difference that the parameter is no longer arc length. As a result we have to have zero instead of -1 in the right hand member of equation 32.1, that is to make A = 0 in the equations 32.5 and 32.7. The equation of the orbit will therefore be (34.1)

$$35.1 \qquad \left(\frac{du}{d\varphi}\right)^2 + u^2 = \alpha + 2Mu^3$$

which will have to be compared with the same equation without the term containing u^3 which is an equation of a straight line and characterizes the propagation of a beam of light on the old theory. In fact, the equation of a straight line whose distance from the origin is 1/p and which is perpendicular to the polar axis is in our coordinates $u = p\cos\varphi$; we have then $du/d\varphi = -p\sin\varphi$ and taking the sum of the squares of the last two expressions we find that they add up to p^2 which we may identify with α. Again the term $2Mu^3$ is very small because the maximum value u can take is the inverse of the minimum value of the distance from the center of the sun, which is the radius of the sun; we treat the problem again as a perturbation problem, that is, compare the required solution to that of the equation without the $2Mu^3$ term. Again we are interested in the change of the angle φ corresponding to a transition between the two extreme values of u. We shall be interested in a beam of light emitted from a star, arriving into our telescope and passing on its way very near to the surface of the sun. The distances of the star and even of the earth from the sun are very large in comparison with the minimum distance, and we shall take them = ∞. The maximum value of u, corresponding to the minimum distance from the sun we shall denote by u_o. Since $du/d\varphi$ changes its sign when the light particle reaches this point it must vanish there so that the left hand side of 35.1 reduces to u_o^2, and we have

$$u_o^2 = \alpha^2 + 2Mu_o^3;$$

we may use u_o instead of α in our equation and write it in the form

$$\left(\frac{du}{d\varphi}\right)^2 + u^2 = u_o^2 - 2Mu_o^3 + 2Mu^3.$$

Solving this for $d\beta$ we find

$$d\varphi = \frac{du}{\sqrt{2M(u^3 - u_o^3) - (u^2 - u_o^2)}}.$$

We introduce a new variable x letting

$$u = u_o \sin x$$

and after the substitution develop according to powers of M and keep only two terms; we thus get for $d\varphi$ approximately

$$\left(1 + Mu_o\frac{1 - \sin^3 x}{1 - \sin^2 x}\right)dx;$$

if we let x change from o to π, u will change from zero to u_o and back to zero, just the change that the inverse distance will experience during the propagation of the light particle. The total change of the angle will then be represented by the integral

$$\int_o^\pi \left\{1 + Mu_o\frac{1 - \sin^3 x}{1 - \sin^2 x}\right\} dx;$$

on the old theory which corresponds to the absence of the term with M in the equation 35.1 we will have to omit the term with M in this integral and the result is π. The approximate result according to the new theory will differ from that by

$$Mu_o\int_o^\pi \frac{1 - \sin^3 x}{1 - \sin^2 x}dx = Mu_o\left[\tan\left(\frac{x}{2} - \frac{x}{4}\right) - \cos x\right]_o^\pi = 4Mu_o$$

The beam of light coming to us from a star will then be deflected by an angle $4M/r_o$ where r_o is its minimum distance from the sun, compared to the old theory, or to the beam as it would go if the sun were absent. If then we observe a star in a certain position on the sky while the sun is far away, and then observe the same star when the sun is near the line of vision; i.e., when the apparent position of the sun is near the

apparent position of the star, this latter position must appear shifted away from the sun approximately by the angle $4M/r_0$. Actual measurements are possible only during an eclipse of the sun, because otherwise the light from the sun drowns out the fainter light from the star, and are beset with difficulties but the results seem to be in favor of the prediction.

36. Shift of Spectral Lines.

We come to the third so-called test of the General Relativity Theory, that is, a case where the predictions of the theory differ from those of older theories by an amount exceeding the error of measurement, thus affording an opportunity to prove or disprove the advantages of the new theory.

In this case again we deal with propagation of light in the gravitational field of the sun, but this time the source is supposed to be on the sun itself, and the observer is on the earth, so that the direction of the beam is that of a radius of the sun; we take this to mean that θ = const. and φ = const. We have then according to 32.5 with A = 0

36.1 $$T^1 \frac{\dot{r}^2}{\eta} - \eta \dot{t}^2 = 0.$$

This gives

$$dr = \pm \eta dt;$$

the double sign corresponds to two possible senses of the beam: from the sun to the earth and from the earth to the sun. The former, in which we are now interested, is characterized by the property that r increases as t increases; the ratio dr/dt must therefore be positive, and since η is positive we must take

36.2 $$dr = \eta dt.$$

The orbit is thus determined by the equations

$$d\theta = 0, \quad d\varphi = 0, \quad dr = \eta dt.$$

But we are interested in the color this time, and color, as we have agreed in Section 16, is proportional to the time component of the momentum vector. As the momentum vector we have to consider the vector of components du_i/dp, where p is a parameter appearing in the equations of Geodesics, the one with respect to which differentiation is denoted by $\cdot$. In order to find this parameter we have to go back to the original equations of geodesics; 32.21 becomes here

$$\ddot{r} - \frac{\eta'}{2\eta}\dot{r}^2 + \frac{\eta\eta'}{2}\dot{t}^2 = 0;$$

equation 36.1 shows that the last two terms cancel leaving us with

$$\ddot{r} = 0.$$

This means that r is a linear function of the parameter,

$$r = ap + b$$

so that

$$d/dp = a.d/dr.$$

The momentum vector $\dot{u}_i$ is here therefore

$$a.\frac{dr}{dr} = a, \quad a.\frac{d\theta}{dr} = 0, \quad a.\frac{d\varphi}{dr} = 0, \quad a.\frac{dt}{dr} = \frac{a}{\eta},$$

the last according to 36.2. What about the value of a? The answer is that it is not and cannot be determined by the foregoing discussion. There are different beams of light which satisfy all the conditions imposed so far; they differ in <u>color</u>, and different colors correspond to different values of a.

Color, according to our definition, is the time components of the momentum vector; i.e., the scalar product of the momentum vector of light and the unit vector in the time direction. If we denote the (contravariant) components of the latter by T^i, the condition that it has time direction will be given by

$$T^1 = T^2 = T^3 = 0$$

and the condition that it is a unit vector - by

$$g_{\alpha\beta}T^\alpha T^\beta = -1,$$

which, taking into account the relations just preceding and the values of the g's becomes

$$\eta(T^4)^2 = 1.$$

The scalar product of the vectors $\dot{u}_i$ and T^i calculated according to the formula $g_{\alpha\beta}\dot{u}_\alpha T^\beta$ becomes

$$\eta . a/\eta . \eta^{-\frac{1}{2}} = a.(1 - 2M/r)^{-\frac{1}{2}}$$

or, if we expand and keep only the first two terms,

$$a(1 + M/r).$$

The color of a beam of light is then not constant along the beam. We shall compare the color as it appears near the surface of the sun, where r is equal to the radius of the sun r_S, and near the surface of the earth, where we may assume $r = \infty$. The frequencies in these two cases will be for a given beam of light proportional to

$$1 + M/r \text{ and } 1;$$

the change in frequency will be proportional to

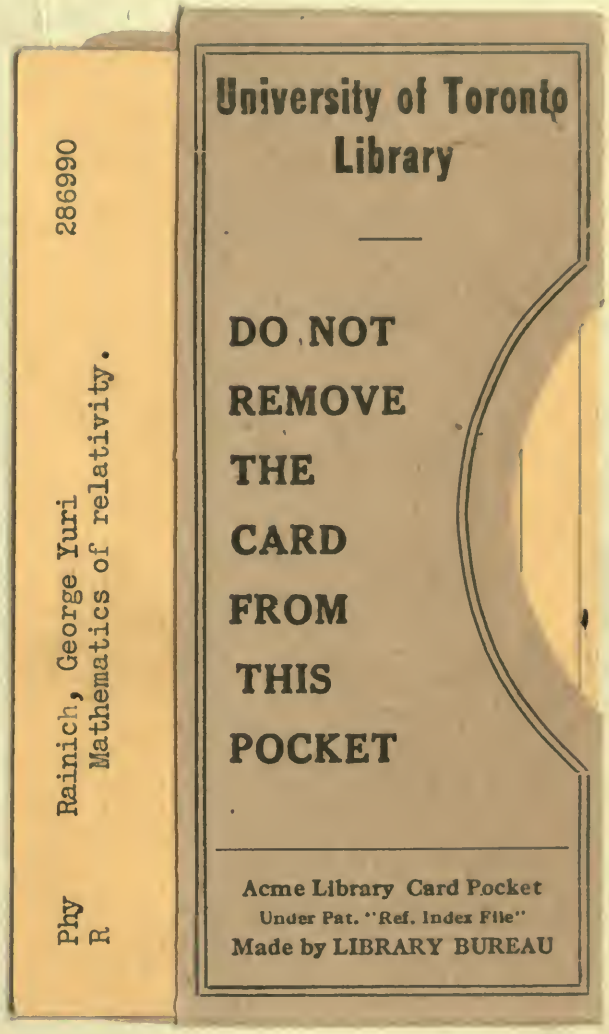
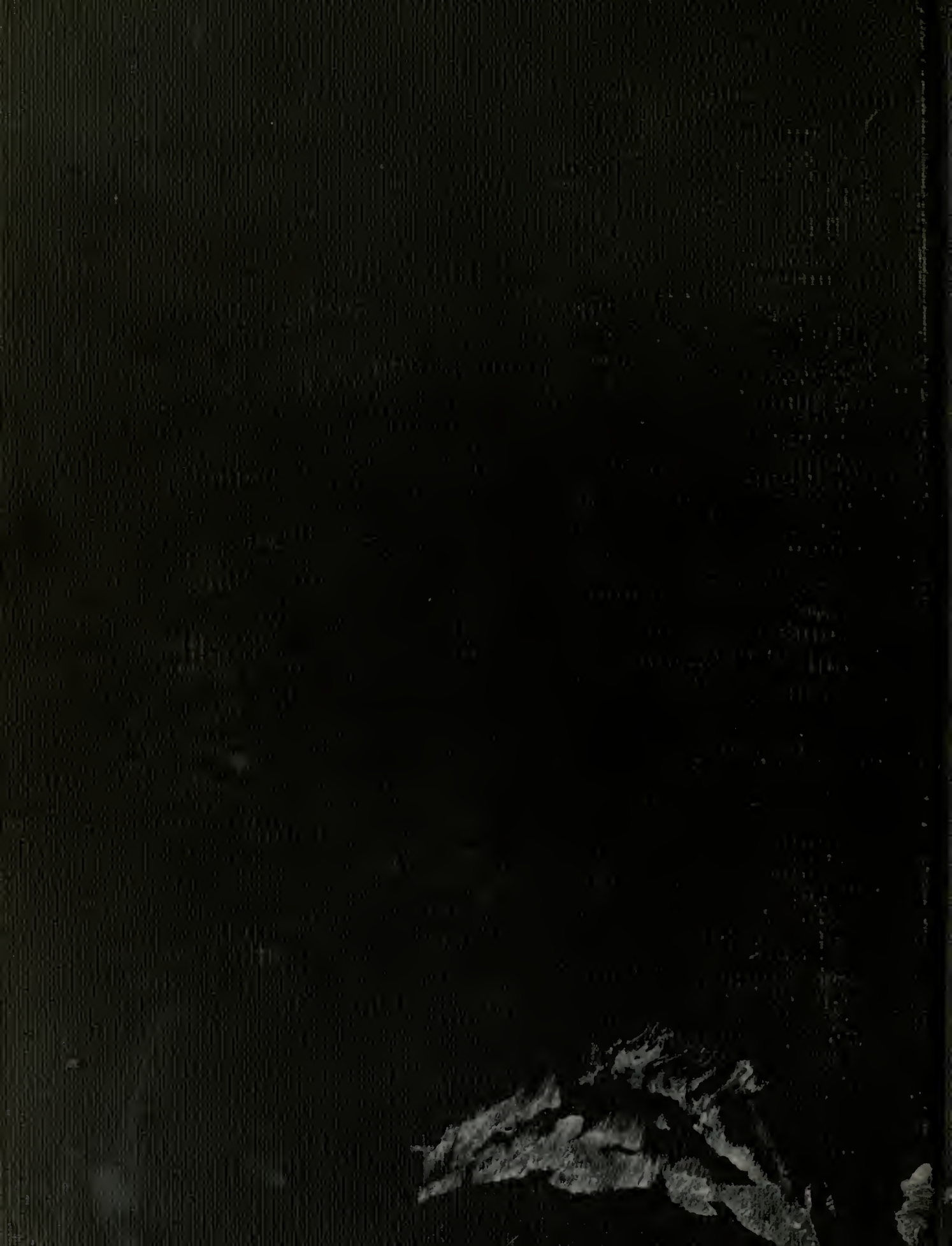
$$(1 + M/r) - 1 = M/r,$$

and this will be also the relative change in frequency.

If now we consider some source of light near the surface of the sun, whose frequency we know, the light emitted by it when it is received at the surface of the earth will have a frequency that is <u>less</u>, the amount of the relative change being given by

$$M/r.$$

If then we compare light coming from a terrestrial source, for instance, emitted by an atom, and light emitted by a corresponding source on the sun, for instance, emitted by an atom of the same kind, we would expect a change of frequency of the amount M/r. Or, if we compare a Solar spectrum with a Terrestrial spectrum, the lines of the former will be shifted toward the red by the amount M/r. This is the prediction of the General Relativity Theory.

Again the experimental evidence seems to favor this prediction.

BINDING LIST APR 1 1933